Tilburg University

Competition and innovation

van der Wiel, H.P.

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
van der Wiel, H. P. (2010). *Competition and innovation: Together a tricky rollercoaster for productivity*. CentER, Center for Economic Research.

Competition and innovation:

Together a tricky rollercoaster for productivity

HENRY VAN DER WIEL

Competition and innovation: Together a tricky rollercoaster for productivity

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Tilburg, op gezag van de rector magnificus, prof. dr. Ph. Eijlander, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de aula van de Universiteit op woensdag 14 april 2010 om 16.15 uur door

HENDRIKUS PIETER VAN DER WIEL

geboren op 28 oktober 1962 te Zwijndrecht.

PROMOTIECOMMISSIE:

PROMOTOR: PROF. DR. J. BOONE

OVERIGE LEDEN: PROF. DR. B. VAN ARK
PROF. DR. E.J. BARTELSMAN
PROF. DR. E. BROUWER
DR. G.M.M. GELAUFF
DR. G. VAN LEEUWEN

Foto: Dragon Khan in Universal Port Aventura (Spanje)

Drukkerij Prisma Print Tilburg

Contents

Preface

For ages, productivity is an interesting topic for economists like me, raising a number of puzzles what drives it to higher levels. This book reflects my work on one of those puzzles: the relationship between competition, innovation and productivity. In the course of time, five people have inspired me to take up this intriguing connection. I am honored that four of them are member of my PhD committee.

Around 1995, I came across the puzzles of productivity for the first time when analyzing differences in performance between Germany and the Netherlands for CPB Netherlands Bureau for Economic Policy Analysis. I contacted Bart van Ark to ask him for more information on all kinds of productivity issues. For years, van Ark is an internationally recognized expert in international comparative studies of economic performance, productivity, and innovation. At that time, he was director of the Groningen Growth and Development Centre, a research group of economists and economic historians examining long-run economic growth and international comparisons of economic performance.

In the mid 1990s, Eric Bartelsman further inspired my curiosity for analyzing productivity issues. Sharing our room, Eric – in those days advisor to the CPB – and I frequently discussed the sources of productivity growth, both from a micro and macro point of view. He particularly convinced me of the importance of using firm level data for productivity analysis since firms are very heterogeneous and may behave differently.

Another important person to mention in this respect is George van Leeuwen. He is an outstanding expert on combining firm level data with econometrics. Together we were for awhile an excellent couple at CPB, publishing a number of interesting papers on the relationship between ICT, innovation and productivity. Without the support by George, I would never have been ranked in the top ten as one of the most cited economist of the top-40 economist list of *Economische Statistische Berichten* in the middle of this decennium.

The fourth person that stimulated me doing this type of research is Philippe Aghion. Combining theory and empirics, he came up with the well documented and often quoted inverted U-shape relationship between competition and innovation in the early 2000s. I once met Philippe in Nice at the end of 2005 presenting my results of this relationship for the Dutch retail trade. At that time, we did not find this inverted U, but the results were only preliminary. Afterwards, Philippe asked me the following in English with his French accent:

"I wanna *neu*". The result you find in this book were we have looked at more Dutch industries. Finally, Jan Boone should be mentioned. Since the end of 2004, Jan and I have been working together on how to measure competition and its relationship with innovation. Jan is a brilliant economist using inventive insights from the Industrial Organization literature. He is the founder of a new competition measure that forms the center of this book. We call it the profit elasticity nowadays, but speaking for myself, people may call it the Boone-indicator. Obviously, it is an inspiration to work with him and I am honored that Jan is my promotor.

The result of these five sources of inspiration is that I have written many papers on productivity topics since the mid 1990s. As a flavor of this output list, it all started with decomposing productivity growth into the contribution of incumbents, entrants and firms that exit the market. Using the growth accounting method at the industry level, I also decomposed Dutch labor productivity growth into proximate causes. Later on, working together with George van Leeuwen, we frequently published on the importance of ICT and innovation for productivity. Lately, I have moved to study the relevance of competition for productivity. A large part of the latter forms the body of this book.

Besides the five people that have inspired me since the mid 1990s, the realization of this book would have never been completed without the help of a lot of people. It would be quite a list and space to provide everybody the credits they deserve. Therefore, I limit myself and I especially want to thank five (former) colleagues from CPB. These are: Harold Creusen, Fred Kuijpers, Bert Minne, Frans Suijker, and Björn Vroomen. Each of you helped me in different ways to analyze and write about *CIP*, as we abbreviated the relationship between competition, innovation and productivity in our jointly projects. Working with all of you was stimulating and a pleasure. Also, the drinks afterwards talking about other subjects than economics, I especially enjoyed very much. I am grateful to my employer, CPB, for giving me the opportunity to work on this thesis, and particularly deputy director George Gelauff as being member of my PhD committee. Finally, Theo Roelandt acts as special opponent in the public defence of my thesis. In fact, Theo was at the cradle of my thesis by asking me to join OCFEB.

To complete this book I am also very much indebted to three colleagues from CentER University of Tilburg and thank them for their great cooperation: Erik Brouwer, Lapo Filistrucchi and Jan van Ours. Three chapters are the result of joint work with them (and Jan

Boone) that we started early 2005 with the financial support by NWO. Erik and I obtained most of the data needed for that research. Moreover, chapter 4 of this book is a joint production with Erik. Lapo and Erik are co-authors of chapter 5, where we come up with new insights into the relationship between competition and innovation. Jan van Ours is co-author of the paper that forms the basis for chapter 2 of this thesis.

Finally, this book is written in Latex. As layman in this field, I greatly thank Berend Hasselman from CPB for helping me with all the technical problems including the layout of this book.

Last but not least, I am very grateful to my family and friends. In particular, I want to thank my parents for educating me other things than economics, and Jan Kijkuit: my paranimf and friend for already more than forty years. Above all, I would like to thank Wietske, my beloved wife, and our three beautiful kids: our daughter and also paranimf Diriëlle, and our two sons Benjamin and Jorik. Your love and understanding were the spirit for me to complete this thesis, but also for the necessary distraction from work. Moreover, although one may call me doctor henceforth, you always give me the right perspective.

Henry van der Wiel

1 Introduction, framework and main results

1.1 Introduction

"Competition is the keen cutting edge of business, always shaving away at costs" (Henry Ford)

"In business, the competition will bite you if you keep running, if you stand still, they will swallow you." (Victor Kiam)

"Innovation is the central issue in economic prosperity" (Michael Porter)

"Just as energy is the basis of life itself, and ideas the source of innovation, so is innovation the vital spark of all human change, improvement and progress" (Ted Levitt)

It seems rather obvious that competition and innovation are important phenomena both in economic theory and in our day-to-day lives, as they both are claimed to generate higher productivity, economic prosperity and more welfare. Buzz words are in this respect: faster, better and cheaper.

People who 15 years ago did not know how to switch on a DOS computer, now surf over the World Wide Web and not only send emails to distant relatives and friends but also call them for free by voip. This is both because of improvements in hardware but also because of innovations that make working with computers easier. Even our bicycle has become highly sophisticated and almost as light as a feather over time. At first glance, the race bike of Maurice Garin - the first winner of the Tour de France in 1903- is seemingly comparable with the bike of Alberto Contador - the latest winner of the Tour de France. However, the materials and technologies of both bikes are very different due to numerous innovations (*e.g.* derailleur gears and fibre). But these innovations lead to higher productivity in terms of average speed in the Tour de France nowadays.

People also experience more competition in all kind of markets: for instance the aviation industry, electricity, telecom, and health markets. Spurred by the European Commission (EC), since the early 1990s, the aviation industry has been liberalized (*e.g.* non discriminatory allowance to airports) and deregulated (*e.g.* less restrictive requirements) with the goal of creating an internal European market for this industry. As consumer, today we can choose between more airliners offering cheap flights to more destinations than ever experienced in the past. Similarly, most telecom markets became liberalized in Europe in the mid 1990s. Before that, people could only choose the national telecom provider and had merely one option to communicate: a fixed telephone. Nowadays, people can choose their own telecom provider and have, due to innovation, numerous options to communicate with each other over the world using cellular phones, email, SMS, twitter, hyves etc. For firms, this means higher productivity since they can organize their production process smarter and their workers can produce more than in the past. For a country, this normally means higher welfare.

This book investigates the intriguing relationship between competition, innovation and productivity, both from a theoretical and empirical perspective.[1] Competition and innovation seem to be indivisibly connected to each other (see *e.g.*, Schumpeter (1934, 1942); Arrow (1962) and Aghion and Howitt (2006)). Competition stimulates innovation by firms, and firms that innovate try to beat their competitors otherwise they will be swallowed by them. Competition as well as innovation are main drivers of productivity growth, but according to recent insights a trade-off may exist between these drivers. In fact, recent findings for the UK suggest that the relationship is shaped like an inverted U (see Aghion et al. (2005)) suggesting that competition is not always positively correlated with innovation. If competition is too intense, it has a negative effect on innovation (and productivity).

This book has two main goals. First, the book sheds more light on how to measure competition on product markets. In that respect, it elaborates on a new competition measure, the profit elasticity, founded and promoted by Jan Boone (see *e.g.*, Boone (2000a) and Boone et al. (2007a)). Chapter 2 and chapter 3 extensively discuss this indicator of competition and explicitly focus on what is meant by 'competition'. The second goal of this book is to ana-

[1] Financial support from NWO (grant numbers 453.03.606 and 472.04.031) is gratefully acknowledged. This thesis is one of the products of CentER for the NWO Research programme *Dynamism in Innovation*. This programme focuses on two issues: (i) innovation, market and hierarchy (ii) innovation and knowledge transfer. The firm level data analysis reported in this book was carried out at the Centre for Policy Related Statistics of Statistics Netherlands. The views expressed here are those of the authors and do not necessarily reflect the policy of Statistics Netherlands.

lyze the relationship between competition, innovation and productivity. Empirical evidence for the relationship between competition and innovation for the Netherlands is hardly available (see Creusen et al. (2006b) for an analysis of the inverted-U curve for only the Dutch retail trade), this book fills this gap by using Dutch (aggregate) firm level data. In fact, chapter 4 explicitly deals with the relation between competition, innovation and productivity at the industry level taking into account other determinants of productivity as well. Chapter 5 examines the link between competition and product innovation at the firm level. Moreover, both chapters also discuss how policy can affect productivity (growth) through competition policy and/or innovation policy. Since the mid 1990s, both competition and innovation are important pillars of Dutch economic policy trying to spur productivity. Examples are longer shop opening hours in 1996 and the founding of the NMa as competition authority in 1998. But if there is a trade off between competition and innovation as Aghion et al. (2005) found for the UK, it challenges researchers and policy makers to come up with the right policy mix. In that case, stimulating both competition and innovation simultaneously might be a *tricky rollercoaster* for enhancing productivity.

The common themes of the chapters in this book are (product market) competition and innovations as main sources of productivity growth. Both sources are interrelated, but competition can also improve productivity without innovation, for example through reducing X-inefficiencies and removing inefficient firms from the market (see for a further discussion below and *e.g.*, Baily et al. (1992); Bartelsman et al. (2003, 2004) and Baldwin and Gu (2006)). The book, therefore, fits in ongoing research for searching for the fundamental drivers of productivity growth (see *e.g.*, Solow (1956); Kendrick (1961) and Jorgenson and Griliches (1967)). The latter is important since productivity growth drives a country's long-run per capita growth rate. In general, productivity growth directly affects the living standards of the population, and thereby the welfare level (see *e.g.*, Canton et al. (2005)).[2] The codified knowledge about the drivers of productivity probably started with the famous publication of Adam Smith 'Wealth of Nations' in 1776.

[2] Note that welfare is not exactly equal to Gross Domestic Product (GDP) per capita because welfare also includes issues like noise and environmental pollution that are not valued in prices or insufficient reflected in GDP.

The current chapter is organized as follows. Section 1.2 introduces the main framework of this thesis. It elaborates on the relationship between competition and innovation from a theoretical and empirical perspective. This section also discusses what we mean by competition and how we want to measure it in the next chapters. Section 1.3 gives the main contributions of this thesis to the literature. Finally, section 1.4 provides a reader's guide by summarizing the main findings of the next chapters in this book.

1.2 Framework of thesis

1.2.1 Economic growth

In general, economic growth as creator of more welfare can be realized in two ways. One way is through employment growth as population growth increases the labor force. Economic growth will be larger if more people participate in the production process.[3] Improvements in productivity are the second way in which economic growth can be enhanced. Those improvements in productivity can be realized by product innovation generating a successful introduction of new products, better quality of products or new services: all with higher value (added) for their users. Also, using new (general purpose) or altered technologies in the production process (*i.e.* process innovation) like investments in information technology (IT) allow people to work smarter and hence raise productivity (see *e.g.*, Jorgenson and Stiroh (2000); Gordon (2000); Van der Wiel (2001a) and Van Ark et al. (2003)). Hence, in the second way to enhance economic growth, more value is produced with given factor inputs, often referred to as total factor productivity (TFP) or multi factor productivity (MFP). TFP can be seen as a measure of an economy's long-term productivity due to innovations including technological changes.

But what should a country or firm do to create productivity growth? And why do competition and innovation particularly matter for this? Looking at growth theories with a helicopter view, two dominant theories exist. The neoclassical growth theory focuses on capital accumulation, while the endogenous growth theory emphasizes knowledge accumulation.

[3] It should be noted that, for instance, a higher retirement age will have a positive, but only temporary effect on economic growth. In the long run, it merely has a level effect on GDP.

Neoclassical growth models assume that productivity growth is exogenous, arising as "manna from heaven". Solow's standard neoclassical growth model predicts that growth through capital accumulations stops (Solow (1956)). Although labor and capital contribute to higher productivity, their contribution dries up in the long run due to diminishing returns to labor and capital. In the end, TFP growth is the sole engine of productivity growth. TFP growth is sometimes referred to as a "measure of our ignorance", since it is measured as a residual within the growth accounting methodology without having a clear explanation what drives this growth measure (see Abramovitz (1956)).[4]

The endogenous growth theory challenges this view of ignorance of the neoclassical theory since the early 1980s. According to this growth theory, deliberate economic behavior and human actions such as investments in innovation and human capital affect long-run economic growth. The endogenous growth theory accounts for long term technological progress and productivity growth without diminishing returns to scale.

In endogenous growth models firms spend resources in response to market opportunities to come up with technological progress. More precisely, the accumulation of knowledge is the underlying source of sustained growth supported by spillovers to other agents of the economy. Human capital (including education, on-the-job training and learning by doing), scientific research, process innovation and product innovation contribute to knowledge accumulation reflected in technological progress. The so called *Schumpeterian growth* (or innovation based growth) models focus on the decisions of firms to conduct R&D in an imperfectly competitive world (see Aghion and Howitt (2006) and the textbox for other endogenous growth models). Due to monopoly power of the successful innovator, the prospect of receiving a profit through better technology gives firms an incentive to invest in innovation.

The debate whether long-run economic growth developments can best be explained from neoclassical or endogenous growth theory is far from being settled. However, the idea that

[4] The growth accounting methodology decomposes economic growth into several input factors and TFP, assuming perfect competition, constant returns to scale and technological change being Hicks-neutral. Oulton (2001); Pilat and Lee (2001); Van der Wiel (2001a) and Van Ark et al. (2003)) are a number of empirical examples applying this method. These TFP-measures as approximation for technological progress are to some extent debatable due to their (neo-classical) assumptions. Moreover, measurement problems might also distort these TFP-measures. These problems are related to, amongst others, how to construct capital services (including services provided by IT) and how to treat R&D expenditures in a growth accounting framework.

Other endogenous growth models

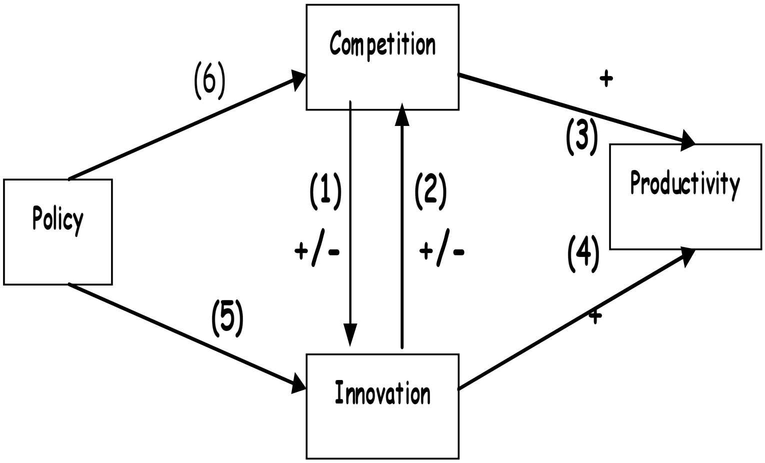
There are at least two other types of endogenous growth models available besides Schumpeterian growth models, mostly referred to as AK-models and Romer's product variety models. However, unlike those growth models, Aghion and Howitt (2006) argue that Schumpeterian growth models produce testable predictions as to how competition (including impact of entry and exit) affects growth.
They claim that the *AK-models* do not say anything on how competition and entry policy affect growth as up to now those models assume perfect competition. Exit is always bad in the *product variety models* as it reduces the economy's GDP by reducing the number of varieties, whereas it has a positive effect on innovation and productivity growth for incumbents but also for the total economy in Schumpeterian theory. Entry is always growth enhancing as it increases product variety. Product variety models 'ignore' the escape competition effect (*i.e.* when competition becomes more intense, it increases the incentive of leaders to innovate). Simply because in those models, entrants innovate whereas the escape competition effect requires that incumbents perform innovations. Moreover, AK-models and product variety models ignore the importance of taking account of the country's or sector's distance to the technological frontier. In this view, expected profits to firms from a successful innovation differ depending on the distance of the industry in question to the technological frontier (*i.e.* the technology giving the highest possible level of output given the inputs), and the threat of entry, measured as the probability of a new firm entering that industry. So, the Schumpeterian growth models seem to fit better in the scope of this book.

education and innovation can contribute to economic growth is now widely accepted among economists and often applied in empirical research (see *e.g.*, Cameron (1998); Griliches (1998); Van der Wiel and Van Leeuwen (2003), and Van der Wiel et al. (2008)). Moreover, in the neoclassical view, firms are homogenous (*i.e.* there is only one type of firm: the representative firm) and operate in a perfectly competitive world. These assumptions are not confirmed in recent empirical evidence. First, firms are heterogeneous (see *e.g.*, Bartelsman and Doms (2000)). They differ in performance due to many underlying sources such as innovative efforts and labor skills. Second, endogenous growth models as the Schumpeterian growth models assume imperfect competition.[5] This assumption better fits the real world.

This book focuses on the relationship between competition, innovation and productivity. We illustrate our main framework with figure 1.1, which pictures the ambiguous relationship between competition and innovation. The arrows present the direction of the causality between the variables. Moreover, the figure also shows the expected impact of one variable on

[5] Note that this is not the case for every endogenous growth model (see box *Other endogenous growth models*).

Figure 1.1 Competition, innovation and productivity



the other. Other important drivers like human capital, may also affect productivity but we primarily examine the former drivers. Further, we ignore the effects of labor market competition on productivity (see *e.g.*, Acemoglu et al. (2006) and Deelen et al. (2006) for reviews of competition on the labor market).[6]

Arrows (1) and (2) refer to the relationship between competition and innovation. Reviewing the literature, this interrelationship is discussed in subsection 1.2.4. Competition is also directly related to productivity (*i.e.* arrow (3)), we consider this link in subsection 1.2.2. Similarly, innovation is connected with productivity (*i.e.* arrow (4)), subsection 1.2.3 reviews this connection. Finally, arrow (5) and arrow (6) show the policy perspectives of innovation measures and competition measures respectively to spur productivity. We examine those measures in subsection 1.2.5.

[6] The empirical evidence related to labor market regulation and productivity is scarce and provides no decisive answer. On the one hand, the effect of employment protection might have a positive impact on productivity as less flexibility of workers might enhance the incentives for employers to invest in specific qualities. On the other hand, rigidities for firms to exit the market and for workers to switch jobs might hamper the process of reallocation of input sources across firms to attain a higher aggregate productivity level.

1.2.2 Competition

Competition is a complex phenomenon and the right way to measure competition is still an unsettled question in the literature. We regard product market competition as the game between firms on product markets in order to maximize their profits. This game is complex as many determinants are involved. We enumerate the following: firms' behavior including their strategic interaction with their competitors, demand of consumers, entry barriers and the prevailing regulation.

We are not aware of a universally accepted definition of the concept of competition. The competition concept that we use in this book captures two ways through which a market (or industry) can become more competitive. First, for given conduct, the number of firms in the market can increase (say, because entry costs fall lowering entry barriers). Second, for given number of firms, competition intensifies if firms' conduct becomes more aggressive (say, firms switch from collusion to more aggressive price competition).

This distinction in the way competition can become more intense is important, because their impact on (traditional) competition indicators differs and influences the interpretation of the development of competition, as we will show below and in the following chapters of this book.

How to measure competition?

Competition authorities, policy makers and economists would like to measure the intensity of competition in a market. For instance, competition authorities are interested whether some particular firm is abusing its market power or whether some firms are colluding, both at the cost of consumers. In addition, policy makers want to know the effectiveness of a policy change that aims to intensify competition in a particular industry.

It is difficult to capture all elements of competition in a single variable. Researchers in the empirical Industrial Organization (IO) literature suggest to measure competition by using indirect indicators. Several measures of competition are used. It seems fair to say that concentration measures, like the Herfindahl index (H), and price cost margins (PCM) are among the most popular ones to assess the degree of competition within a market. However, these measures can sometimes provide a wrong impression of the competition intensity in a market. From a theoretical point of view both measures have severe drawbacks, in the sense that they can incorrectly show an increase in competition, when in fact competition has declined (see *e.g.*, Tirole (1988)).

H measures the concentration of firms' market shares in a market. In antitrust, concentration measures are important both in merger cases and in abuse cases (see, for instance, Bishop and Walker (2002)). The idea behind concentration ratios (like H) is that a low level of concentration is seen as fierce competition because it includes many firms operating on a market. A rise in H is then interpreted as a decrease in competition. Only, if more firms enter the market due to lower entry barriers, H rightly reflects this rise in competition as this indicator falls due to less market concentration. But as noted by Tirole (1988), this is not always the case. If intense competition driven by more aggressive interaction removes inefficient firms from the market (*i.e.* selection effect) or reallocates revenues and consequently market shares from inefficient to efficient firms (*i.e.* reallocation effect), market concentration rises. Hence, H will go up suggesting less intense competition whereas actually more intense competition is the reason for a rise in the level of concentration.

The PCM, also known as 'Lerner index', is widely used as a measure of competition in papers like Nickell (1996), Nevo (2001) and Aghion et al. (2005), but it is not a perfect measure of competition either. The PCM refers directly to the firm's ability to set its prices above its marginal costs. Generally, if there are many competitors on a market with a low level of demand, then competition forces the firms to reduce prices until marginal costs. At the other extreme, a monopolist experiences no competition at all and thus can set a high price to maximize its profits. In the range from no competition to perfect competition, the PCM falls. The problem is, however, that a high PCM may point to a monopoly or an industry with high fixed costs, but it could also be the case that some firms are very efficient with high mark ups. Theoretical papers like Stiglitz (1987), Bulow and Klemperer (2002) and Amir (2002) present models where more intense competition leads to higher instead of lower PCM. Similarly as for H, more intense competition due to more aggressive interaction may have a positive effect on the (industry) PCM, as efficient firms (with high PCM) gain market shares at the expense of inefficient firms (with low PCM). The rise in PCM at the industry level due to this reallocation effect incorrectly suggests that competition became less vigorous. But also the selection effect may distort the relationship between PCM and competition. As more intense competition removes inefficient firms with low PCM from the market, consequently the average PCM of the remaining firms will go up.

This book introduces a new measure of competition that is more robust from both a theoretical and an empirical point of view than those traditional measures. We call this measure the profit elasticity (PE). It relates profits and efficiency at the firm level. PE is estimated for a market (or industry) and is defined as the percentage fall in profits due to a percentage increase in a cost variable that captures the efficiency of a firm. In all markets, an increase in costs per unit of output reduces a firm's profits. However, in a more competitive market, the same percentage increase in those costs will lead to a bigger fall in profits. The underlying intuition is that in more competitive markets, firms are punished more harshly (in terms of profits) for being inefficient.

Unlike H and PCM, PE is based on an econometric specification. To estimate PE, one requires firm level data from an unbalanced panel linking efficiency to profits (see chapters 2 and 3 for further discussion). PE cannot be measured from aggregate data like data from the National Accounts of Statistical Offices. However, if firm level data are available, then the data requirements are the same as for PCM as similar variables (*i.e.* sales and a cost measure) are needed for both estimating PE and the calculation of PCM. In that sense, it is easier to calculate H at this aggregation level as this indicator needs only sales to measure the market shares, but the interpretation with respect to the development of competition is not always clear as discussed. In this book, we claim that PE can be measured from an unbalanced panel. In contrast, the observations of all firms active on a market including firms producing in foreign countries are needed for measuring H correctly. Although PCM is the only indicator from the three competition indicators concerned that can be calculated at the industry level using aggregate data, the above-mentioned reallocation - and selection effects can bias the relationship between PCM and competition. To overcome this problem, the correct calculation of PCM requires a balanced panel, particularly for concentrated industries were the reallocation effect can be substantial.

Link to productivity

In line with the thoughts of Henry Ford, more intense competition increases productivity in the following ways (see arrow (3) in figure 1.1).

First, competitive pressure stimulates firms to operate efficiently by, for instance, 'cutting the fat out' of their organizations. An increase in competition may force firms to achieve the highest level of efficiency in production and management, given available technologies.

That is, increasing competition may reduce various forms of X-inefficiency like managerial slack and bureaucratic inertia, and subsequently enhances productivity in the market (see *e.g.*, Nickell (1996); Aghion et al. (1999); Lever and Nieuwenhuijsen (1999) and Disney et al. (2000)).[7] This is called productive efficiency. *Productive efficiency* refers to the efficiency in the use of inputs to produce some (given) quantity of output, or stated otherwise the extent to which total costs to produce the quantity of output is minimized. In that regard, more intense competition weeds out inefficient firms from the market, increasing aggregate productivity.

Second, fiercer competition also brings prices in line with marginal costs, lowering the rents of producers and increasing consumer surplus. Vigorous competition may therefore result in more productivity as resources and output are allocated to their most productive use in the economy. This is called *allocative efficiency*, it refers to the match of supply and demand such that resources are allocated in the most efficient use.

Static efficiency builds on the two above-mentioned concepts: productive and allocative efficiency. Static efficiency is the extent in which total surplus is maximized in the short run.[8] Put differently, a market is statically efficient if the combined welfare of consumers and producers is maximized, while production takes place using the current technology and its inputs in the most optimal combination.

1.2.3 Innovation

In contrast to the concept of competition, the (endogenous growth) literature is more coherent on what innovation exactly is. New and/or improved technologies, better products or services are all aspects of innovation that can enhance productivity.[9] Hence, innovation is related to the word ‘new’: something that was not available before. Yet, the implementation in (endogenous) growth theory is not straightforward, because innovation has many dimensions. Moreover, besides with competition, innovation also interacts with other drivers of productivity such as human capital. To come up with new ideas is also related to knowl-

[7] A few studies like Scharfstein (1988) and Martin (1993) claim the opposite from a theoretical perspective. These studies argue that competition leads to an increase in managerial slack, and hence lowers productivity.

[8] Notice that in this way we ignore the political choice in weighing consumer and producer surplus. In fact, we consider consumer and producer surplus as equally important.

[9] In the old Neoclassical framework, there is no room for innovative efforts as productivity growth is exogenous and no R&D resources are needed.

edge and skills (see *e.g.*, Romer (1990) and Grossman and Helpman (1991)). Knowledge – driven by education, training and experience – generates those new things, and these things stimulate TFP-growth. As mentioned before, we limit our analysis to the (direct) relationship between innovation and productivity, and neglect other determinants and interactions except competition that contribute to higher productivity.

The general idea in the endogenous growth theory is that (codified) knowledge has the characteristics of a public good (see *e.g.*, Romer (1990)). That is new ideas, designs and blueprints can be non-rival and non-excludable. Hence, the inventor cannot prevent using his idea by others reducing the incentive to innovate. However, protection of this property right provides firms the incentive to innovate as the prospect of (monopoly) profits stimulates firms to innovate. Consequently, the innovating firm can enter the market and replaces the incumbent (*i.e.* Schumpeterian creative destruction). The speed of the innovation process determines economic growth in the end.

How to measure innovation?

How to measure innovation (*i.e.* product - and process innovation) particularly in the context of the interaction with competition? Two indicators that are well known and frequently used in theory and in practice, are: (i) R&D-expenditures and (ii) patents (or patent applications).

R&D-expenditures seems to be an appropriate indicator for measuring the efforts on producing 'new ideas'. Investments in R&D are widely used in endogenous growth models (see for instance, Romer (1990), Grossman and Helpman (1991) and Aghion and Howitt (1999)). But, R&D is an input measure, it does not tell anything about the effectiveness of the innovation process. Conditional on other drivers, a low R&D ratio (*i.e.* R&D-expenditures as percentage of an output measures like sales) may go hand in hand with high productivity when the R&D process is very efficient or productive: few R&D inputs generate relatively large amounts of sales because of successful introduction of new products. Moreover, nowadays, it is recognized that R&D does not capture all the innovative efforts of firms, industries or countries. Other (input) aspects are important as well for measuring the innovative effort in the right way. For example, training, number of R&D workers, non-technological innovations like marketing and organization may also contribute to higher productivity. Further, the entrance of new firms to the market are often not seen or measured as innovation expenditures in innovation surveys by Statistical Offices, but their entrance is often the result of an invention of something new.

Although this measure focuses on the output of the innovation process, patents has shortcomings as well. For instance, not every innovative firm applies for a patent due to, amongst others, high costs of application and keeping the innovation secret. Further, it takes a long time between the application for a patent and its impact on sales due to this innovation.

Due to the availability of surveys like Community Innovation Survey (CIS), innovative efforts can nowadays be measured at the firm level in various ways (see *e.g.*, Brouwer (1997), Kleinknecht et al. (2002) and Van Leeuwen (2009)). Alternative indicators in this respect are: (i) sales of products new to the firm or new to the market, (ii) innovation expenditures. The advantage of the first indicator is that it is an output measure: the sales generated by innovation. But this measure has also disadvantages. For instance, it focuses only on new products and the term new to market is a bit vague as well. The definition of innovation expenditures is much wider than the one for R&D, because the former also consists of costs such as costs of patent application and wages of R&D personnel. Still, innovation expenditures are an input measure. Below we employ both types of alternative measures in our analysis.

Link to productivity

Many studies have investigated the impact of R&D on productivity growth (see *e.g.*, Cameron (1998) and Griliches (1998) for overviews). In general, the empirical literature points to a positive effect of innovation on productivity at the firm level without giving an unambiguous result of the size of this effect (see arrow (4) in figure 1.1).

Cohen and Levinthal (1989) and Griffith et al. (2004) provide empirical evidence that R&D may even have "two faces" with respect to productivity. First, firms conduct R&D in order to generate own innovations for their products or production process. This first face – the innovation part – reflects the direct effect of R&D on productivity growth of firms. Second, firms may use their own R&D in order to absorb knowledge and to adopt innovations from other firms. To some extent, followers may reap benefits from cheap or costless imitation, for instance by adopting codified knowledge of frontier firms that is freely available (no licenses) and that can be applied without any adjustments. But in order to reap all benefits from imitation they may also apply some own R&D to enhance their absorptive capacity, particularly to regenerate and/or adapt tacit knowledge in order to implement innovations in firm's own products and process. Therefore, the second face of R&D – the imitation part – refers to the benefits of knowledge spillovers for productivity. Note that with the potential for imitation, the social rate of return on innovative R&D is larger than the private rate of return. More precisely, an innovating firm cannot appropriate all the benefits of other imitating firms that may accrue from its innovation. Firms may even abstain from innovation if their costs of innovation exceed their private (expected) benefits, notwithstanding the possibility that the social benefits may be higher than the costs of innovation.

Innovation is directly related to dynamic efficiency. *Dynamic efficiency* denotes the extent to which the present value of a (future) stream of total surpluses can be maximized over time (long enough to allow for investments in product and process innovation). Total welfare over a longer period of time, and thus dynamic efficiency, can be improved via product and process innovation (see *e.g.*, Baumol (2003)). Better products (new products or higher product quality) will increase consumers' willingness to pay and entail an upward shift in consumer demand. Additionally, improved or new production techniques, which reduce firms' (marginal) production costs, entail a downward shift of the supply curve.

1.2.4 Relationship competition and innovation

Taking into account the interplay between product market competition and innovation, economic theory does not predict how competition affects productivity and economic growth in the longer run (see arrows (1) and (2) in figure 1.1). Whether or not competition raises innovation is an ongoing debate and a challenging research topic since Schumpeter's remarks in his two famous books, dividing the theoretical strands into two camps. The first strand consists of those that argue that competition can be bad for innovation (see Schumpeter (1942)). The second strand states that competition can be good for innovation (see Schumpeter (1934)). Since those two books of Schumpeter, many theoretical and empirical studies have tried to settle this relationship without an unambiguous answer (yet).

Competition bad for innovation

According to Schumpeter (1942) fiercer competition generates less R&D, reducing the rate of innovation and hence economic growth. The intuition is that if the expectation of higher profits drives innovation then an increase in competition will reduce (post) innovation if it results in lower gains from innovations. Firms need monopoly profits to provide the incentive to innovate. Using a Schumpeterian endogenous growth model, Aghion and Howitt (1992) showed that an increase in product market competition has a negative effect on productivity growth by reducing the monopoly rents that reward innovation (see also Romer (1990) and Grossman and Helpman (1991)). The empirical studies that support this negative correlation are limited, but Hamberg (1964); Mansfield (1964) and Kraft (1989) are examples.

Competition good for innovation

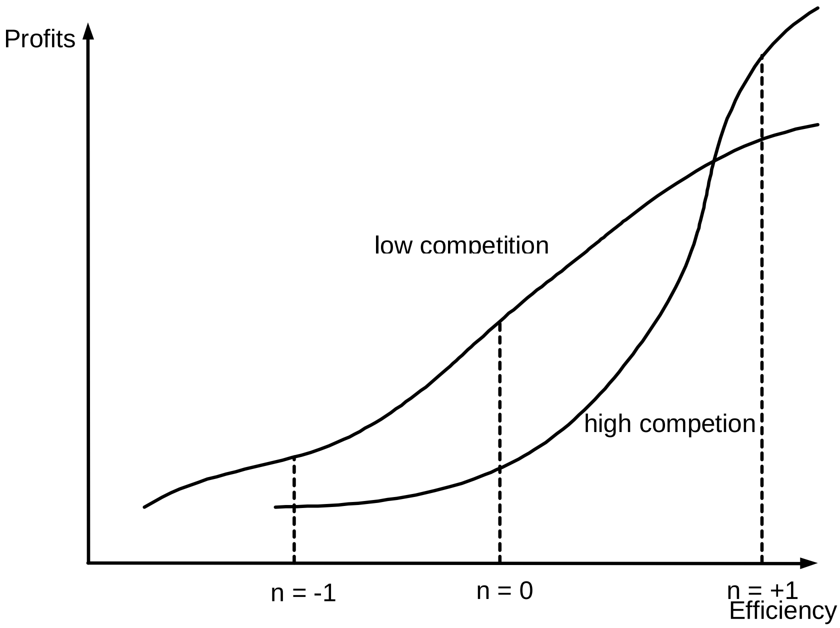
Studies from Schumpeter (1934); Arrow (1962); Scherer (1980) and Porter (1990) express the view that competition is good for innovation. In this strand, it is thought that competition stimulates incumbents to innovate otherwise the firm is forced to leave the market and the potential entrant will win the race. This entrant will win this race if the replacement effect (Arrow (1962)) for the incumbent is stronger than its efficiency effect (see below). When innovating the incumbent monopolist replaces her own profits while the potential entrant has no pre profits to replace at all. Again, Aghion and Howitt (1999) showed these mechanisms in an endogenous growth model. Competition encourages innovation and economic growth, because it reduces incumbent's pre-innovation profits more than it lowers its post innovation profits. The empirical evidence for this second strand is larger than for the first strand. We refer to studies like Geroski (1990); Blundell et al. (1995, 1999); Nickell (1996) and Carlin et al. (2004) that find a positive relationship between competition and innovation (or productivity).

Recent literature: An inverse U relation?

Aghion et al. (2005) tries to capture the main effects from both strands and comes up with an inverse U relation between competition and innovation: both a positive and negative effect of competition on innovation may arise depending on the initial level of competition.

There are two main effects driving this inverted U-shape. First, in a very competitive market the profits of a laggard do not depend much on how far he is behind the leaders: his profits are almost zero anyway. While in a market with soft competition, efficiency does affect the profit level for inefficient firms. This is illustrated in figure 1.2 which considers a duopoly where firms can be either leveled (both have same efficiency level) or unleveled (the leader is one ($n = +1$) step ahead of the follower ($n = -1$)).

Figure 1.2 Competition, profits and efficiency



Intensifying competition reduces the incentive for laggards (n = –1) to catch up with the leaders (assuming that innovation is of the 'step by step' variety and not of the 'leapfrogging' sort). This can be seen in figure 1.2 as the profit gain from moving from n = –1 to n = 0 is smaller in case of high competition compared to the case of low competition. Hence, we have the effect here that an increase in competition reduces R&D effort. Second, in a very competitive market it pays off handsomely if you can outperform your opponents. As competition intensifies in figure 1.2 the difference between being ahead (n = +1) and being level (n = 0) increases. Thus, making the market more competitive increases the incentive of the leaders to innovate and move ahead of their opponents. This is the effect where more intense competition leads to more innovation.

By introducing a composition effect, Aghion et al. (2005) find that the latter effect dominates for low levels of (initial) competition while the former effect dominates for high levels of competition. Initially when competition is low, industries are most often leveled. Then, an increase in competition stimulates innovation due to the escape effect. If competition increases further, industries become more frequently unleveled, whereas the chance that it becomes leveled again reduces as for laggards it is increasingly difficult and costly to catch up. Consequently, as the innovation rate is lower in unleveled situations, beyond some level of competition, innovation will decline, generating the inverted U.

This maximizing level of competition may differ across industries. It depends on the ability of (lagging) firms to absorb and to imitate the innovation of the leading firm. This pace of imitation is affected by how firms can keep their innovation secret for their competitors. In that respect, the way intellectual property rights (IPR) like patents are organized matters. Less stringent protection of patents may have a positive effect on innovation (see Boone and Van Damme (2004)).

This inverted U curve between competition and innovation can also occur in another way, because there can be a trade-off between process and product innovation when competition is raised. At the industry level, this may generate an inverted U-curve if one relates the total innovation expenditures (*i.e.* the sum of process and product outlays) to the extent of competition. The reasoning is as follows.

Boone (2000b) shows that a rise in competition may raise industry wide efficiency through more process innovation. In contrast, this may reduce product variety or the number of products introduced to the market: less product innovation. Why? Inefficient firms are forced to leave the market because of a selection effect and lower costs of opponents (higher efficiency level from process innovation) due to higher competition. This reduces the product variety or product innovations in an industry. Moreover, more competition reduces (ex post) profits of most firms and makes it less attractive to introduce a new product. Hence, a trade off may occur between process and product innovations at the aggregate level.

There are, however, two ways that may overturn this trade off between process and product innovation. First, firms could also escape competition by investments in product differentiation, and in doing so creating their own niches within this industry (or market). This may change the market structure of this industry and the intensity of (measured) competition.[10] Second, lower (expected) profits due to more competitive pressure could act as a wake up call for managers. To avoid bankruptcy, managers have to look for new products than can

[10] Whether or not the intensity of competition actually changes depends on the easiness of firms to enter this niche. Moreover, as long as there is no abuse of market power or institutional barriers in those niches, and this product diversification reflects consumer preferences, there is no market failure (see below for further discussion).

generate additional profits. Hence, although process innovation is applied industry wide, innovation expenditures regarding product innovation might go up as well in that particular industry.

This brings us to another question: what is the identity of the innovator? This is a question where Schumpeter changed his mind over time. According to Schumpeter (1934), often referred to as Mark I, innovating firms are new small firms and they challenge the incumbent firms (by so-called 'creative destruction'). Those firms can more easily introduce fundamental breakthroughs as they are better equipped to step into new technological trajectories and have the flexibility of overcoming organizational inertia. According to Schumpeter (1942), often referred to as Mark II, the large established firms are responsible for technological progress. Those incumbents will defend their leading market position against potential entrants by investing in R&D. As incumbents have more to lose, their incentive to raise R&D investments is stronger than that of potential entrants. The incumbent firm avoids substituting high monopoly profits by lower oligopoly profits, while the potential entrant realizes the oligopoly profits at best.

Here the main intuition can be summarized using two effects. On the one hand, there is the *Arrow replacement effect*: when an incumbent innovates, he replaces (cannibalizes) his old product with a new one. Hence, the incentive to innovate is a profit difference, while for the entrant (or a small existing firm with little current profits) the incentive to innovate is the profit level earned after innovating as his pre-profits are zero by definition. This effect makes it more likely that small or entering firms innovate instead of big incumbents. On the other hand, there is the *efficiency effect*: when an entrant innovates, he will still be faced with an important competitor (unless the innovation is so drastic that the incumbent disappears, which does not happen often). Hence, the entrant will earn 'only' duopoly profits, whereas the incumbent when innovating earns monopoly profits. In addition, when the monopolist does not innovate, he loses his current monopoly profits. This also gives an incentive for the monopolist to innovate. The efficiency effect works in the direction of the big incumbent firms innovating instead of small or entering firms.

Boone (2001) links these effects to the ones above on the relation between competition and innovation. He argues that in very competitive industries the efficiency effect tends to dominate. Hence, in a very competitive industry we should expect a leader to increase his dominance. In weakly competitive industries, the Arrow replacement effect dominates and we should expect small and entering firms to innovate. Here we have the reversed causality effect from innovation on industry structure. So, this structure may change as a result of firms' innovation decisions, and may also change the character of competition as well as the pressure to innovate. Such feedback mechanism may cause an endogeneity problem that researchers have to take into account in examining the relationship between competition and innovation.

Link to productivity

As discussed, Aghion et al. (2005) and Aghion and Howitt (2006) show that there can be an inverted U-curve between competition and innovation. But, what is the impact of this relationship on productivity? Using the study of Aghion et al. (2006) it can be shown that the effect of entry on aggregate productivity is always positive, at least in theory. From the model of Aghion et al. (2006) one can deduct that a decline in innovation expenditures (of incumbents) in an industry go hand in hand with higher aggregate productivity. The reasoning is as follows. After intensifying competition, the least efficient domestic firm has no incentive anymore to imitate or to innovate due to the large productivity gap (see *e.g.*, Kocsis et al. (2009)). Consequently, the total innovation expenditures of that industry decline. Yet its aggregate productivity rises. The reason is the entry of a foreign leader with the highest productivity level in that particular industry. That foreign firm replaces the least efficient domestic firm, enhancing aggregate productivity at the industry level.

1.2.5 Policy perspective

Dutch policy intends to foster productivity by stimulating both innovation and competition (see arrows (5) and (6) respectively in figure 1.1). As innovation and competition are important determinants for higher productivity or for higher welfare, it is rather logical that those drivers are key variables for policy aiming for higher welfare. But when should government intervene in markets?

From an economic approach, welfare is not optimal if market failures or government failures are present. If market failures exist, government may consider intervening if and only if government failures are smaller than those market failures: the social benefits of intervention exceed the costs. Here, we particularly consider two potential market failures that are closely related with innovation: abuse of market power, and externalities related to knowledge spillovers and business stealing effects.[11] Abuse of market power by incumbents may occur if incumbents can prevent entry or if institutional barriers protect them for (the threat of) entry. The results are too high prices for consumers and consequently a welfare loss. R&D or innovation activities generate knowledge spillover effects creating benefits that cannot be fully appropriated by the inventor. Hence, the inventor (or firm) has fewer incentives to innovate, while the social rate of return is larger than the private rate of return.[12] This also leads to a welfare loss. Hereafter, we give a number of examples of EU and Dutch policy measures that focus on either competition or innovation.

With respect to competition, policy makers took various measures to raise the competitive pressure in product markets during the 1990s and early 2000s. In general, competition policy tries to limit the abuse of market power and assesses market concentrations that could decrease competition and welfare. International policy examples are the removal of barriers to the internal market of the European Union (EU) in 1992, the policy agenda set by the Lisbon European Council in 2000 and WTO-agreements. On top of that, Dutch policy makers renewed the Competition Act ('Mededingingswet') in 1998, and the NMa, the Dutch competition authority, was founded. The efforts by the NMa to break cartels and to punish collusion are ways in which policy can make the interaction between firms more aggressive and overcome abuse of market power of incumbents. In the period in question, Dutch policy makers also reformed regulations in the so-called MDW-operation (In Dutch: Marktwerking, Deregulering and Wetgevingskwaliteit) to stimulate competition in specific industries, and they privatized sectors like telecommunication. Examples in that respect are: abolishing minimum prices, extending shop-opening hours and liberalizing taxi permits (see *e.g.*, Creusen et al. (2006b)).

Concerning innovation, both the EC and the Dutch government aim to stimulate innovation via subsidies or IPR. Subsidies are given based on the idea that an innovative firm cannot fully internalize the benefits of its innovation due to knowledge spillovers. For instance, the Seventh Framework Programme for Research and Technological Development (FP7) of the

[11] Other market failures are information asymmetry (for instance on capital markets) and network externalities (including economies of scale, for instance on telecom markets).

[12] There can, however, be a negative externality from innovation. Firms that innovate do not take into account the social costs due to making other innovations obsolete. This is called the business stealing effect.

EC is the EU's main instrument for funding research in Europe. The best-known example for the Netherlands is the WBSO (in Dutch: "Wet Bevordering Speur- en Ontwikkelingswerk"), where the subsidy is in the form of a reduction of payroll tax and social security contributions of R&D workers. Other ways to spur innovation is by protecting innovation with IPR like patents giving the inventor a temporary monopoly power ex post to internalize the gains of its innovation.

So far, our discussion focuses on generating either high static efficiency through stimulating competition or high dynamic efficiency through stimulating innovation, all else equal. However, knowledge is lacking when it comes to the dynamic effects of innovation and competition policies taken together. Main questions that then arise: (i) Where should government focus on? (ii) What if the inverse U-relationship is present? Figure 1.1 points at the challenge for policy makers who aim at boosting productivity growth and try to find the right balance between competition and innovation policy. For policy, the following three issues can serve as guidance to start with.

First, it is productivity as an indicator of welfare that should be the main target for policy and not, for instance, the intensity of competition or the amount of innovation expenditures in a particular market. As illustration, entry per se should not be the goal of economic policies, but affecting the size of entry costs can be a potential instrument to improve productivity through competition (*e.g.* reducing institutional entry barriers). Low entry rates may correlate with high productivity, because the threat of entry is what really matters for the behavior of incumbents. Further, as already discussed, according to the theory of Aghion et al. (2006) lower innovation expenditures (of incumbents) due to more intense competition can go hand in hand with higher productivity. If there are no substantial entry and exit barriers, the entry of a foreign leader will replace a less innovative and productive domestic firm. Hence, the focus on innovation expenditures as policy goal instead of welfare (or productivity) can be misleading.

Second, in principle policy instruments should tackle the core problem of the market failure at its roots. If incumbents abuse their market power through too high mark ups, collusion or deterring entry, then competition legislation should prevent this. Similarly, well designed IPR's and innovation subsidies should address production externalities as market failures directly.

Third, and finally, an ex ante cost-benefit analysis (CBA) needs to precede each intervention of policy. A CBA-analysis assesses the costs and benefits of the policy measure considered. It also takes into account alternative measures to determine whether the proposed measure is legitimate, effective and efficient. Such analysis should also include the costs of intervention in terms of policy -, enforcement - and transition costs (*e.g.* reallocation costs of labor and bankruptcy).

Suppose policy makers know that for an industry competition and innovation are positively correlated, then they have three options to stimulate productivity if this is needed because of market failures: (i) use an innovation measure and (ii) use a competition measure, (iii) use both. We argue that the choice between these options is not directly straightforward without profound analysis with input from studies like the ones reported in this book.

When more intense competition stimulates innovation, then competition improves static and dynamic efficiency. So, in such case the arguments to liberalize and deregulate industries become stronger. But, it also poses the question whether new innovation measures are still needed as more competition already positively affects productivity via more innovations. To answer, this question policy needs a CBA that should come up with answers to questions like: What are the consequences for the amount of knowledge spillover within and across industries due to stronger competition in a particular industry? Are firms in that industry more inclined to keep their information secret, and hence the social benefits for the economy at large will be lower?

The implication for policy gets even more complicated if the inverse U-relationship between competition and innovation is actually present, and it is likely that an industry is in the downward sloping part of this relationship. Then what the most effective and efficient policy measure is, is not directly clearcut. In that case, a trade-off between static and dynamic efficiency exists and the aim of innovation policy and competition policy can be at odds with each other. Both high static and high dynamic efficiency are not jointly attainable in this situation. Policies that encourage competition (for example through lowering entry barriers) may then be detrimental for innovation (*i.e.* high static efficiency and low dynamic efficiency: relatively low (consumer) prices but limited technological progress). Similarly, stimulating innovation by patents might enhance dynamic efficiency, but may create a deadweight loss if the inventor abuses his market power (low static efficiency with high prices).[13]

Again, in such case a CBA is needed for policy. Here, the choice for policy makers is between high dynamic efficiency but low static efficiency or the opposite. The CBA has to take into account the benefits and costs of both situations, its uncertainties, risks and the (social) preferences of current and future generations (see *e.g.*, Canton (2002)).

Competition not always conducive to welfare

More competition may reduce welfare through the so called business stealing effect (see *e.g.*, Dixit and Stiglitz (1977); Mankiw and Whinston (1986); Aghion and Howitt (1992) and Aghion and Howitt (1999)). When an innovative firm enters the market, this firm does not take into account the social costs of making already existing innovations obsolete. This so called business stealing effect works in the direction of too many firms entering the market from a welfare perspective. On the other hand, there might be too few firms that enter from a welfare perspective. This is related to appropriability effect, as firms cannot perfectly price-discriminate between consumers and other clients. In that case, those firms are not able to appropriate the full consumer surplus and too few firms will enter the market (*i.e.* rent spillovers). Moreover, innovations often create (intertemporal) knowledge spillover effects generating benefits that accrue beyond the succeeding innovation that the innovative firm cannot internalize. So, whether more competition has a positive or negative effect on welfare depends on which of the three effects dominates.

Wrapping up, when the dynamic effects of innovation and competition are taken together, the implications for policy are not immediately clear. In that respect, empirical research on the relationship between competition and innovation is already useful for policy makers, as it helps to gain insights in this link although without providing the full answer. For example, whether such trade-off between competition and innovation is present in practice. Our results show that such trade off is present in the Netherlands, but that it is ex post. Competition increases the innovation incentives, at least for product innovation. However, within an industry the firms that have successfully introduced new products are the one that face less intense competition.

[13] Note that this trade off forms the core of the debate on the scale of optimal protection by IPR. IPR creates ex post market power for the inventor. A higher value of IPR stimulates innovation ex ante but generates a higher deadweight loss as well ex post. Hence, a trade off occurs between stimulating innovation and reducing the deadweight loss in the design of IPR. Then, the main issue is where to stop with protection or the duration of protection. The issue is also related to the social welfare dilemma between knowledge creation and knowledge diffusion (see CPB (2002)).

1.3 Main contribution of thesis

State of the art research on the empirical relation between competition and innovation as Aghion et al. (2005) did for the UK has not been done for Dutch industries yet. Hence, we do not know whether there exists an inverted U curve for the Netherlands, and which industries are beyond the maximal innovation/competition level. It should be noticed that the latter cannot be directly learned from Aghion et al. (2005) as they analyze firms and not industries.[14]

For the Netherlands, the relation between competition and innovation is especially interesting since for many years its performance on productivity growth is relatively low in an international perspective, particularly compared to the US, pushing the Netherlands back in their top-ranking with regard to the level of productivity (see e.g., Van der Wiel (2001a); Gelauff et al. (2004) and Van der Wiel et al. (2008)). Some people claim this is due to a lack of competition in many Dutch industries. The problem, however, is that with the current lack of knowledge about the connection between competition and innovation we cannot meaningfully inform Dutch policy on what to do.

This book fills this gap of knowledge and it also contributes to the vast amount of literature searching for the drivers of productivity (growth) in a number of ways.

First, it shows how one can measure competition using firm level data. It is not always clear what different studies mean by 'competition' and how policy can affect competition in these cases. Aghion et al. (2005) use the PCM as a measure of competition. We argue that this is not a robust measure from a theoretical and an empirical point of view and that the PE is a better competition measure in particular situations than the PCM when panel data of firms is available (see chapters 2 and 4). Using firm level data, we show that the PCM is not always monotone in competition in case of more aggressive interaction between firms, whereas the PE is.[15] The more concentrated the industries are the more likely the development of PCM cannot be interpreted as the development in competition. We present empirical evidence that the percentage of industries were both measures (*i.e.* PE and PCM) do not point in

[14] To illustrate, firm A of industry X can be located in the upward sloping part of the figure, while firm B of the same industry can be in the downward sloping part.

[15] Concentration ratios like H are even more problematic in case of aggressive interaction.

the same direction in terms of competition development can be sizeable in the case of the Netherlands, and therefore should not be ignored. Hence, one should be careful to use PCM as a measure of competition in empirical research, particularly in concentrated industries where the reallocation and selection effects can be considerable when competition changes. Then the reallocation effects might be substantial because efficient firms gain market shares at the cost of inefficient firms. Likewise, the selection effects are sizeable if inefficient firms are removed from the market and efficient firms increase their market shares. Both effects move PCM in the wrong direction with respect to competition. Chapter 3 shows how to measure PE in practice and that this measure is robust in a number of ways.

Second, the book examines the relationship between competition and innovation (see chapters 4 and 5). Using firm level data as well as industry data, we replicate the Aghion et al. (2005) analysis of the inverse U relation for Dutch industries. This exercise is already informative for researchers and policy makers, as it helps to gain insights into the link between competition and innovation. Both the theoretical and empirical literature provide an ambiguous answer how competition may affect innovation. In that respect, we take into account that there might be a link from innovation back to competition (*i.e.* reverse causality). Product innovation may reduce the intensity of competition for instance by making products less close substitutes. We find evidence for an inverted U curve between competition and innovation using industry level data, but the implications for policy differs from the one from Aghion et al. (2005). We show that this trade off might be ex post (see chapter 5). More intense competition stimulates an industry to innovate more. But within the industry the firms that have successfully introduced new products are the one that face less intense competition.

Additionally, the book applies better innovation measures for investigating the connection between competition and innovation than used by Aghion et al. (2005). They use patent data (weighted by citations), however, it is well known that patents are an incomplete measure of innovation covering only a small part of all innovations. Many innovations are not patented by firms but simply kept secret. Our data allows us to identify such innovations as well, giving a broader picture on actual innovation activities of firms.

The third contribution of the book is that it relates both competition and innovation to productivity (see chapter 4). Neither competition nor more innovation should be the main goal for policy as these are only "intermediate variables" that may improve productivity. Hence, higher productivity (or more welfare) should be one of the main targets for policy. We, therefore, consider the impact of both competition and innovation on productivity (growth) because of the productivity problem in the Netherlands mentioned above. The book addresses the issue whether fiercer competition or direct stimulation of innovation (or R&D) raise the productivity performance in the Netherlands. We show that competition is a more promising channel to stimulate productivity through innovation than giving innovation subsidies to firms. But, here we should also take into account the evidence for an inverted U curve: competition can be too fierce. This may eventually have a negative impact on productivity. However, our estimation results indicate that this occurs at levels of competition that are far beyond levels observed in general. Hence, in general, more competition is always better for (product) innovation.

The final contribution of this book to the literature is that it underlines the importance of using firm level data in this type of research, and the importance of taking into account industries beyond manufacturing industries (see chapters 2, 4 and 5). The availability of firm level data allows us to consider heterogeneity of firms. Differences in productivity performance can be due to various reasons related to the underlying sources of productivity. For instance, we refer to differences in applied technology, management quality, labor skills, and innovative efforts. Moreover, we can control for different institutions as we have data across industries. Furthermore, we link firm level data to industry level data to take account of the variance of a variable next to its mean. This is a rather new way to analyze economic behavior with aggregate data. Finally, the study of Aghion et al. (2005) and its finding of the inverted U curve is only based on data for the manufacturing industry. This book examines other industries too, like services. We show that this distinction matters in terms of level of competition, but also whether or not an inverted U shape between competition and innovation exists.

1.4 Structure of thesis

This last part of the current chapter gives the structure of this dissertation and it therefore produces a reader's guide by summarizing the main findings of this book. Two general remarks need to be taken into account before reading this book.

First, the rest of the book consists of four chapters.[16] Although competition and innovation are the main topics of this thesis, each chapter is a separate study on a particular subject and therefore readable in itself. However, it implies that once and awhile we repeat ourselves in other chapters. For instance, in describing the data sources used and by reviewing the theoretical literature on the relationship between competition and innovation.

Second, the data set used in this book is not completely the same for all chapters due to time sequence of the research. Chapter 2 uses an older data set with fewer observations than that in the rest of the book.

In chapter 2 (*Measuring Competition*) we discuss and apply a new measure of competition: the elasticity of a firm's profits with respect to its efficiency level captured by its average variable costs (AVC). A higher value of this profit elasticity (PE) signals more intense competition. Using firm level data from the 'Produktie Statistieken' (PS) for approximately 250 Dutch markets, we compare PE with two most popular competition measures: the price cost margin (PCM) and the Herfindahl-index (H). Competition can become more intense in two ways: (i) lower entry barriers, (ii) more aggressive conduct of firms. The first way gives no problem. The three indicators are correctly picking up the change in competition. Next, we show that PE and PCM are often correctly picking up the second way as well, but H is at odds. However, PCM is not always right. It tends to misrepresent the development of competition over time in markets with few firms and high concentration, *i.e.* in markets with high relevance for competition policy and regulation. So, just when it is needed the most PCM fails whereas PE does not. From this, we conclude that PE is a more reliable measure of competition in case firm level data is available.

Chapter 3 (*Robustness of Profit Elasticity*) analyzes the robustness of the estimation results of PE using fixed effect (FE) estimation techniques. This chapter provides a guide for researchers how to measure PE in practice. It assesses what the effect on the estimated PE is under a range of other conditions to find out whether PE is a robust measure for analyzing the developments of competition. These conditions include alternative model specifications, different econometric estimation techniques, and the impact of measurement errors and selec-

[16] Chapter 6 is the Dutch summary of this thesis.

tivity issues. For doing so, we employ a data set containing more than 320,000 observations over the period 1993-2006 based on PS information of about 121,000 individual firms in the Netherlands from 154 industries at the 3-digit SIC-level. The results of this chapter can be summarized as follows. First, tests for the functional form hint towards a loglinear specification, making the interpretation of PE as being an elasticity easier. Second, the idea of the relationship between profits and AVC is to a large extent robust to different ways in which PE can be estimated in econometrics. The results of PE based on FE are significantly correlated with the results of pooled OLS, random effect, and first difference estimation procedures. Nonetheless, our preference for the FE-estimation technique is supported by using F-tests and Hausman test (Hausman (1978)). Third, we explore a couple of sensitivity tests to assess the robustness of our FE-model when taking into account potential measurement and selectivity issues in the panel data set. Again, the results for PE are robust as the correlations with our basic specification are highly significant.

Chapter 4 (*Competition and innovation: Pushing productivity up or down?*) examines the relationship between competition, innovation and productivity for the Netherlands. We use industry level data as well as moments from firm level data for more than 150 three (or sometimes four) digit SIC-industries covering almost the whole Dutch economy over the period 1996-2006. We match innovation data from Community Innovation Survey (CIS) with accounting data from PS to link innovative activities with performance at the industry level. Starting from a production function, we include ideas from the endogenous growth theory, where competition and innovation may both affect total factor productivity (TFP). Moreover, looking at the convergence literature, we add the distance to the frontier (*i.e.* highest productivity level) to our econometric specification as an additional determinant of productivity growth. We consider the endogeneity problem with respect to competition and innovation by using instruments via the Generalized Method of Moment (GMM) estimation technique. The main findings of our analysis can be summarized as follows. First, we find strong evidence for a positive impact of competition on TFP at the industry level. Competition directly increases TFP by reducing X-inefficiencies and removing inefficient firms from markets. Second, this chapter finds evidence that there may exist an inverted U-curve between competition and innovation for the Netherlands, at least for manufacturing industries. This corresponds with findings of Aghion et al. (2005). Moreover, our indicator for innovation subsidies turns out to be not significant for innovation. Apparently, competition is the most important determinant of innovation and this determinant is not always conducive to innovation expenditures. When competition becomes too fierce it may have a negative effect on productivity via lower innovation expenditures. However, combining all estimation results, it turns out that this is at levels of competition that are far beyond levels observed in general. Therefore, intensifying competition is a promising option for policy makers to raise productivity in the Netherlands given the current innovation policy. Third, we find no evidence for a negative feedback mechanism from innovation back to competition for the aggregate economy. In the sense that too high levels of innovation may reduce the competition intensity. For the manufacturing industry, we do find indications for such a feedback, but this occurs beyond high levels of innovation intensity. Lastly, as indicator for competition, we use the PE in this study. To test the robustness of this indicator, we also applied the PCM as alternative indicator. The latter turns out to be not significant in any equation concerning productivity or innovation, making the PE an interesting measure in productivity research to proceed on.

Finally, chapter 5 *(Product innovation reduces competition intensity)* focuses on the relationship between competition and innovation at the firm level. This chapter particularly examines the effect of product differentiation related to making products less close substitutes, and hence making markets less competitive. The idea is that more intense competition leads industries to innovate more. However, firms innovate to reduce competition, making competition endogenous (reverse causality). Hence within an industry, successful innovators of new products are the ones that face less intense competition after product innovation. We identify these effects of product innovation reducing the competition intensity using Dutch firm level data covering large parts of the Dutch economy over the period 1993-2006. In fact, we obtain panel data from matching two data sources from Statistics Netherlands – PS and CIS – to measure competition and product innovation. We use PE as competition measure. As indicator for product innovation, we exploit two types indicators. The first type of innovation indicator captures whether a firm has applied for a patent. Since usually quite some time elapses between applying for a patent and introducing new products based on that patent, we conjecture that this variable is not affected by the endogeneity problem. That is, a firm that has applied for a patent is not (yet) able to use the patent to differentiate its products from its competitors, and hence affect the level of competition. The other type of indicator that we use is whether the firm has recently introduced new products in the market. If product differentiation plays a role, we expect to see an effect for innovation variables of this type on competition. Summarizing our main findings, we come up with an alternative explanation for the negative correlation between competition and innovation, and hence for the trade off between static and dynamic efficiency. We claim, however, that the policy implication is the opposite: more competition is always better for (product) innovation in industries! However, firms that have innovated manage (ex post) to reduce the competition intensity that they face. Thus we find ex post a trade off between dynamic and static efficiency. Indeed, once we look inside industries (by using industry or firm fixed effects), the correlation between competition and innovation remains positive for the variable based on patent applications but turns negative for variables capturing new products introduced in the market. That is, within a market (or industry) the firms that introduce new products are the ones that face relatively little competition. We interpret this as innovating firms differentiating themselves from competitors and in this way reducing the competitive pressure that they face in their market.

1.5 Epilogue

This book comes up with findings that contribute to our understanding of productivity (growth), at least for the Netherlands. Nonetheless, a number of potential determinants deprived from the theoretical literature turns out to be less or not relevant at all in practice. For example, we did not find evidence for the relevance of the distance to the frontier as explanation for higher productivity. This could be due to incomplete theoretical insights. On the other hand, measurement errors could be the reason as well, underlining the importance of up-to-date measurement methods by Statistical Offices. In that respect, we refer to human capital and its measurement. The endogenous growth theory sees human capital as one of the main determinants of productivity growth, while currently statistical observations are most often insufficient to take them into account in productivity analyzes. Further, better instrumental variables for both competition and innovation would be welcome to cope with the endogeneity problem in the relationship between competition and innovation. Having data from NMa cases or relevant policy interventions in innovation issues makes this type of research presented in this book even more policy relevant.

The search for the drivers of productivity started centuries ago, but it has to go on in future since many questions are not yet solved. A part of the tricky rollercoaster has been passed through, however our brains (read knowledge) are yet not able to fully collect our thoughts what really happens in the world.

2 Measuring Competition

2.1 Introduction

In the empirical Industrial Organization (IO) literature, several measures of competition are used.[17] It seems fair to say that concentration measures, like the Herfindahl index (H), and price cost margins (PCM) are among the most popular ones.[18] However, from a theoretical point of view both measures have severe drawbacks (see below and, for instance, Tirole (1988)).

This chapter introduces a new measure of competition that is more robust both from a theoretical and an empirical point of view than those 'traditional measures'. We call this new measure the profit elasticity (PE).[19] PE is measured for a product market and is defined as the percentage fall in profits due to a percentage increase in (marginal) costs. In all markets, an increase in costs per unit of output reduces a firm's profits. However, in a more competitive market, the same percentage increase in (marginal) costs will lead to a bigger fall in profits. The underlying intuition is that in more competitive markets, firms are punished more harshly (in terms of profits) for being inefficient.

This chapter argues that PE is in some cases a better competition measure than PCM and H. One way to make this point would be to show that PE corresponds more closely to the definition of competition. Unfortunately, we are not aware of a definition of the concept of *competition*.[20] However, we think it is not controversial to distinguish the following two ways in which competition can be intensified in a market: (i) more firms in a market due to a fall in entry barriers and (ii) more aggressive conduct by incumbent firms. We analyze the implications of both these ways to intensify competition on the measures H, PCM and PE.

[17] This chapter is based on Boone et al. (2007a,b). We thank Harold Creusen, Lapo Filistrucchi and Free Huizinga for useful comments and suggestions on earlier versions.

[18] In antitrust, concentration measures are important both in merger cases and in abuse cases (see, for instance, Bishop and Walker (2002)). In the empirical literature, PCM is used as a measure of competition in papers like Aghion et al. (2005), Nevo (2001) and Nickell (1996).

[19] The measure is based on theoretical research in Boone (2000a) and Boone (2008).

[20] Of course, "perfect competition" is defined. However, there is no generally accepted way in which "more intense competition" in an oligopoly context is defined.

More firms entering a market tends to lower concentration in this market. Hence, more intense competition due to more entry is correctly picked up by a concentration measure like H. The problem with concentration measures as indicators of competition is, however, that a switch to more aggressive behavior by firms (*e.g.* because a competition authority detects and abolishes a cartel that manages to raise price and divide the market between participants) forces inefficient firms out of the market (selection effect of competition). This raises concentration, but should (clearly) not be interpreted as a fall in competition. Also more aggressive conduct by firms tends to raise the market shares of efficient firms at the expense of inefficient firms. Such a reallocation (of market share) effect raises H as well. We show that in our data set this reallocation effect (and selection effect) can dominate, leading to a (seemingly inconsistent) positive correlation between H and PE.

When PCM is used as a measure of competition in the empirical literature, it is usually calculated as aggregate (variable) profits over aggregate revenues for a particular market. This 'market PCM' can also be written as the weighted average of firms' PCM's where the weights are given by firms' market shares (see, for instance, Nickell (1996)).[21] An increase in competition tends to reduce firms' PCM's.[22] If competition is intensified due to a fall in entry barriers, PCM falls: correctly indicating more intense competition. However, if competition is intensified due to more aggressive conduct, the reallocation effect of market shares can counteract this effect. In particular, an increase in competition raises the market share (and therefore the weight in the calculation of the market average PCM) of efficient firms with high PCM's. Hence the weight of efficient firms (with high PCM) goes up which can raise the market (or aggregate) PCM. This is the main problem with the market PCM as a measure of competition that we focus on: an increase in competition due to more aggressive conduct can actually raise the market PCM due to the reallocation effect. We identify this effect in the data.

We also find evidence suggesting another problem with PCM can play a role as well. If

[21] When firm level data is not available this is the only market PCM one can calculate. This is an advantage of the market PCM compared to PE which does need firm level data to be estimated. However, firm level data is becoming more widely available nowadays. Moreover, by comparing PE and PCM in our firm level dataset, we indicate in which markets it is (relatively) safe to use PCM as a competition measure.

[22] This is actually not always the case as shown in papers by Amir and Lambson (2000), Bulow and Klemperer (2002) and Stiglitz (1987). There an increase in competition (through an increase in the number of firms in the market) can actually raise some firms' PCM's. We do not address this problem in this chapter.

a firm's costs fall over time, its PCM tends to go up. Such an increase in PCM should not be interpreted as a fall in competition. Indeed, conditional on a firm's costs, a high PCM indicates market power. But, conditional on price, a high PCM reflects efficiency.

Using Dutch firm level data for 250 markets over the period 1993-2002, we show that PE picks up the effects of competition in an intuitive way and, in fact, in a way similar to PCM. We consider the correlation between PE and market characteristics like labor income share, import penetration, average productivity levels of the firms etc. These correlations are comparable to the results found for PCM. But the results for H differ considerably from the correlations found for PE and PCM. From this we conclude that H is less suitable as competition measure than PE and PCM.

Although PCM and PE look similar, they are not identical. When considering the change in competition, we find the following. In situations where the reallocation effect is strong, PCM and PE may differ in the direction of the development of competition (one suggesting that competition went up from one year to the next, the other that it went down). This happens in concentrated markets with high H and few firms. Theory presented below then suggests that in these cases PCM fails while PE still is a consistent measure of competition. Note that the effect of concentration implies that PCM and PE deviate in markets that are particularly interesting for a competition authority: in highly concentrated markets an increase in PCM can be caused by more intense competition. This strengthens the point made by Fisher (1987) that PCM is not a good measure of monopoly power.

The next section discusses the related literature. In section 2.3, we use simulations to illustrate the features of the PE measure. Further, we show circumstances under which PE and PCM deviate. Section 2.4 describes the data on competition measures and shows some key statistics. Section 2.5 analyzes PE, PCM and H in more detail. It shows that PE and PCM are correlated in a similar way with market characteristics like labor income share, import penetration etc. Then we show that an increase in competition intensity tends to increase PCM (incorrectly suggesting softer competition) if H is high and the reallocation effect is large. Section 2.6 concludes.

2.2 Related literature

There are numerous papers using measures of competition in the empirical IO literature. The use of concentration as a measure of competition goes back to the structure-conduct-performance framework. High concentration is then seen as a signal of weak competition which leads to high prices and high price cost margins. See, for instance, Scherer and Ross (1990) for an overview. Although, H as representative of the concentration rate indicators is easy to calculate if firm level data is available, its relation with competition is, however, not always straightforward as the discussion above showed.

The PCM also has a long tradition as a measure of competition. Some papers (like Aghion et al. (2005) and Nickell (1996)) calculate it directly as the profits-sales ratio. Others, first estimate demand and cost functions and then calculate the optimal PCM for each firm under an assumption on the relevant competitive model for the firms in the sector. Examples here include Berry et al. (1995), Hausman et al. (1994) and Nevo (2001). By comparing a direct estimate of the PCM (like the profits-sales ratio) with the PCM predicted under different competitive regimes, one can identify which competitive regime applies in a sector. This method has been criticized by Corts (1999) who shows how the transitory nature of demand shocks leads to overestimation of competition intensity. We do not take a stand on the issue of how the PCM should be estimated. However, because we want to give an overview of how the competition measures vary over markets and time, the direct way of calculating the PCM has obvious advantages. To illustrate, with the direct method we do not need to gather additional information for all the markets in our sample, like cost instruments, product characteristics and instruments for consumers' taste parameters (such as demographic variables).[23]

Hall (1988) developed a method to test for a positive PCM without actually calculating it directly. The idea is that under constant returns to scale and perfect competition, the Solow residual is not affected by instrumental variables like military spending and the oil price (Klette (1999) generalized this method by allowing for increasing returns to scale). We do

[23] In the estimation of the PE measure, similar issues arise. In particular, one can choose a structural method to derive the demand and cost curves and then from these curves calculate the profit elasticity. For the reasons mentioned, we do not use such a structural model in this chapter.

not use this method for two reasons. First, the Hall-method tests whether there is either perfect competition or market power. In the markets where there is market power (most of the markets in our sample), the method does not provide a degree of market power. Second, for this method convincing instrumental variables are needed which we do not have for all the markets in our sample. Roeger (1995) adapts Hall's method by combining primal and dual estimates of the Solow residual. This allows him to estimate mark ups without the use of instrumental variables. However, for Roeger's method data is needed on the capital stock and the rental rate of capital. Constructing a capital stock is rather complicated with firm level data, therefore we do not use this method to estimate PCM. Moreover, we are particularly interested in changes in the competition measures over time. Roeger's method only provides an average PCM over time.

The PCM is often interpreted in a normative way: lower PCM is "better" in the sense that it is associated with higher welfare. Although this is true in a very simple model, in general there is no clear relation between PCM and welfare.[24] Further, as pointed out by Fisher (1987) the profits-sales ratio is not a good measure of monopoly power because the user cost of capital is hard to measure.[25] The PE measure avoids this problem for two reasons. First, admittedly a bit trivial, there is no simple benchmark for PE (like the –supposedly optimal– zero benchmark for PCM). As shown by Boone (2003) the welfare maximizing value of PE depends on the characteristics of a market, like the cost structure in the market and consumers' tastes. Second, it is not so much the levels of profits and costs that are important for PE. The crucial issue is how a change in efficiency captured by (marginal) costs causes a change in profits. To the extent that capital costs are fixed costs, we actually do not need to take them into account (although we will show that high capital costs are associated with less intense competition). A related point here is that empirical evidence is mounting which shows that more intense competition leads to more innovation and higher efficiency (see, among others, Aghion et al. (2005), Klette (1999), Nickell (1996) and Porter (1990)). If less intense competition leads to higher costs due to X-inefficiency or lack of innovations to reduce costs, PCM is reduced. This causes an overestimation of competition

[24] See, for instance, Mankiw and Whinston (1986) and Amir (2002) for examples where lower PCM does not imply higher welfare.

[25] Fisher and McGowan (1983) give a related criticism on the use of accounting rates of return to infer market power.

using PCM. Again because PE does not focus on profits and costs *levels*, it avoids this pitfall. This is not to say that PE is robust to imperfections in the data. However, compared to PCM data problems are not worse and may even be partly alleviated by PE.

The PE introduced here is reminiscent of the measure based on factor price elasticities used by Panzar and Rosse (1987). In particular, they show that the sum of the factor price elasticities of a monopolist's revenue, denoted by ψ, must be nonpositive: $\psi \leq 0$. Hence, if $\psi > 0$ for a firm, it is not a monopolist. If $\psi = 1$, the firms in the sector are in a long-run competitive equilibrium. In a monopolistic competition outcome one finds $\psi \leq 1$. The statistic ψ is derived as a test for monopoly. However, using ψ as a measure of competition has two main drawbacks. First, if $\psi \leq 0$, we actually do not learn anything, except that the sector is not in a long run competitive equilibrium. A negative sum of elasticities is consistent with both monopoly and oligopoly. In the oligopoly model used, one is the upperbound on ψ. There is no sense in which ψ closer to one implies a more competitive sector. Second, to calculate ψ one needs information on factor prices. This is usually harder to come by than information on revenue and costs. Moreover, we have no information on factor prices in our data set.

Finally, whereas we are mainly interested in competition in terms of aggressiveness of conduct, Bresnahan and Reiss (1990) and Bresnahan and Reiss (1991) focus on competition in terms of entry. They focus on geographically isolated markets (for the same product) to establish the relation between the size of the market and the number of firms in the market. This indirectly gives information on firms' conduct. As we are interested in the developments of competition measures economy-wide over time, we do not use this relatively time-consuming method to derive information on market power.

2.3 Model

We use the following notion of competition. In a more competitive market, firms are punished more harshly in terms of profits for being inefficient. In fact, PE estimates a relation between firms' profits and efficiency captured by (marginal) costs in a market. This section presents the theory underlying this measure. The starting point is that there are two ways in which competition can be intensified. First, competition becomes more intense as the number of firms in a market increases (for given conduct) due to a fall in entry costs. Second, competition becomes more intense as firms' conduct become more aggressive due to for example changes in consumer preferences. Using simulations we show that the competition measures PCM and H work well in the former case but not in the latter, this particularly involves H. We argue that PE picks up both forms of changes in competition correctly.

2.3.1 Introduction of profit elasticity

In any IO model, the relation between firm i's profit π_i and marginal cost level c_i (that captures efficiency) is downward sloping. Higher marginal costs c_i imply –for given price p_i– a lower margin per unit of output sold. Further, higher marginal costs tend to lead to higher prices, which reduces the amount of output x_i sold. Roughly speaking, we use the following specification of this relationship

$$\ln(\pi_i) = \alpha - \beta \ln(c_i). \tag{2.1}$$

With this linear specification between $\ln(\pi_i)$ and $\ln(c_i)$, which can be viewed as a first order Taylor approximation, the slope β can be interpreted as an elasticity. It indicates the percentage fall in profits due to a one percent increase in marginal costs. We call β the profit elasticity, PE.

To interpret PE, first consider a simple monopoly model where the firm faces a constant elasticity demand function $x = p^{-\varepsilon}$ where x denotes output and p the price charged. We assume $\varepsilon > 1$ and constant marginal costs $c > 0$. This monopolist chooses output level x to solve

$$\max_x \{x^{\frac{\varepsilon-1}{\varepsilon}} - cx\}$$

It is routine to verify that output is given by $x = \left(\frac{\varepsilon-1}{\varepsilon c}\right)^{\varepsilon}$ and profits by $\pi = \frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}} c^{-(\varepsilon-1)}$. Hence we find $\ln(\pi) = \alpha - \beta \ln(c)$ with $\alpha = \ln\left(\frac{(\varepsilon-1)^{\varepsilon-1}}{\varepsilon^{\varepsilon}}\right)$ and $\beta = \varepsilon - 1 > 0$. Hence in this case the linear relation between $\ln(\pi)$ and $\ln(c)$ fits the model perfectly. Higher β here implies that the monopolist faces a more elastic demand curve, which indeed limits the monopolist's market power. In general the fit will not be perfect and equation (2.1) should then be interpreted as a linear approximation. Further, if the firm is not a monopolist but faces competitors, then ε is interpreted as the firm's own price elasticity or the elasticity of its residual demand curve (which exceeds (in absolute value) the market demand elasticity).

To get intuition for the case with more than one firm, consider the following standard Cournot model. There is a market where each firm i produces only one symmetrically differentiated product, faces an inverse demand curve of the form

$$p(x_i, x_{-i}) = a - bx_i - d\sum_{j\neq i} x_j, \tag{2.2}$$

and has constant marginal costs c_i. This linear demand curve implies that the elasticity is not constant and hence equation (2.1) is not a perfect fit. The parameter a captures the size of the market, the parameter b captures the market elasticity of demand and the parameter d captures the extent to which consumers see the different products in a market as close substitutes for each other. Firm i chooses output x_i to solve

$$\max_{x\geq 0}\{(a - bx - d\sum_{j\neq i} x_j)x - c_i x\}, \tag{2.3}$$

where we assume that $a > c_i > 0$ and $0 < d \leq b$. The first order condition for a Cournot Nash equilibrium can be written as

$$a - 2bx_i - d\sum_{j\neq i} x_j - c_i = 0. \tag{2.4}$$

Assuming N firms produce positive output levels, one can solve the N first order conditions (2.4). This yields

$$x\left(c_i\right) = \frac{\left(\frac{2b}{d} - 1\right)a - \left(\frac{2b}{d} + N - 1\right)c_i + \sum_{j=1}^{N} c_j}{\left(2b + d(N-1)\right)\left(\frac{2b}{d} - 1\right)}. \tag{2.5}$$

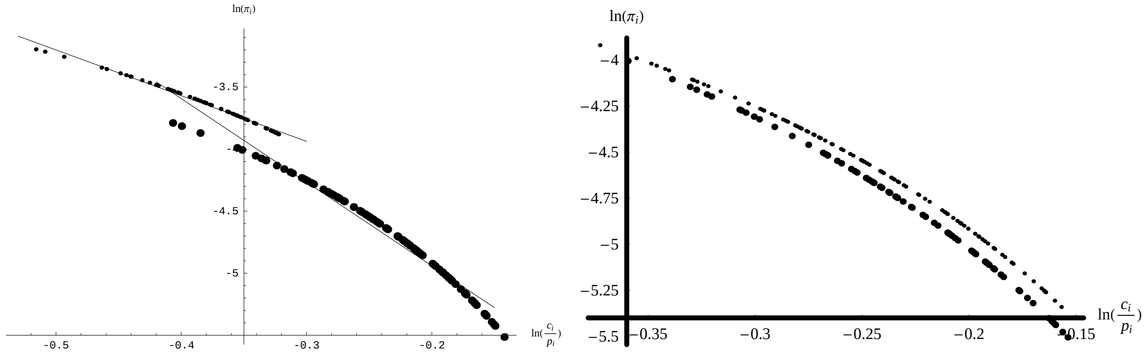
We define a firm's variable profits as $\pi\left(c_i\right) = (a - bx\left(c_i\right) - d\sum_{j\neq i} x\left(c_j\right))x\left(c_i\right) - c_i x\left(c_i\right)$. These are variable profits in the sense that they do not include the fixed cost f.

A firm with marginal costs c_i enters the market if and only if $\pi\left(c_i\right) \geq f$ in equilibrium. This fixes the number of firms N that enter in equilibrium where we assume that more efficient firms enter first. Since we cannot directly observe c_i in the data, we approximate marginal costs with average variable costs defined as $\frac{c_i x_i}{p_i x_i}$. Hence, the relation we are interested in is between $\ln(\pi_i)$ and $\ln\left(\frac{c_i}{p_i}\right)$. In general, our estimates for PE based on $\frac{c_i x_i}{p_i x_i}$ will be lower than based on c_i due to dividing it by p_i.

To compare the behavior of the three competition measures, we use as starting point the standard Cournot model for the case with $a = 40, b = 30, d = 20, f = 0.004$. Further, we draw randomly cost levels c_i for 110 firms out of a lognormal distribution with mean 0.7 and standard deviation 0.08. Now, we change competition in two ways: (i) we change the entry cost and (ii) we consider the effects of more aggressive interaction between firms. We begin with the first way.

Figure 2.1 Relation between $\ln(\pi_i)$ and $\ln\left(\frac{c_i}{p_i}\right)$

$f = 0.02$ **(small dots) and** $f = 0.004$ **(large dots)** $d = 20$ **(small dots) and** $d = 30$ **(large dots)**



The left part of figure 2.1 presents a simulation with a change in the entry cost. The comparison presented in the figure is between a situation with high entry cost ($f = 0.02$) and low entry costs ($f = 0.004$). The relationship is steeper, PE is higher and competition more intense with low entry costs than with high entry costs, as one would expect. In this case, PCM and H are lower with the lower entry cost (see textbox *Definitions of PCM and H* for their definitions). In particular, PCM falls from 0.32 to 0.22 and H falls from 0.016 to 0.010. Hence, all three measures clearly indicate that lower entry barriers lead to more intense competition. This is true more generally: reductions in entry barriers leading to more firms in the market and therefore more intense competition are correctly picked up by all three measures. Note that with the higher entry cost, firms' profits tend to be higher to cover this entry cost. Hence the observations with $f = 0.02$ feature higher values on the vertical ($\ln(\pi_i)$) axis. Second, prices will be higher with higher entry cost (fewer firms in the market). For given draws of c_i, the values of c_i/p_i shift to the left.

The second way to intensify competition is by more aggressive interaction between firms. In this case, we increase competition by making goods closer substitutes: raising d from 20 to

Definitions of PCM and H

To be able to analyze the intensity of competition at the industry level, firm level PCM's are aggregated into a weighted average industry PCM in the following way

$$PCM(\theta) = \sum_{i \in I} s(n_i, \theta) pcm(n_i, \theta) = \sum_{i \in I} s(n_i, \theta) \frac{R(n_i, \theta) - C(n_i, \theta)}{R(n_i, \theta)} \tag{2.6}$$

where I denotes the set of firms in the market, n_i firm i's efficiency level, θ captures the degree of competition in i's industry, R and C present respectively revenues and variable cost, and $s(.)$ denotes firm i's market share

$$s(n_i, \theta) = \frac{R(n_i, \theta)}{\sum_{j \in I} R(n_j, \theta)} \tag{2.7}$$

The advantage of defining PCM in the following way, is that no firm level data is needed to calculate (industry) $PCM(\theta)$, since it can also be derived from industry aggregate information on revenues and costs

$$PCM(\theta) = \frac{\sum_{i \in I} R(n_i, \theta) - \sum_{i \in I} C(n_i, \theta)}{\sum_{i \in I} R(n_i, \theta)}$$

A change in competition intensity θ affects PCM in the following way

$$\frac{dPCM}{d\theta} = \sum_{i \in I} \frac{ds(n_i, \theta)}{d\theta} pcm(n_i, \theta) + s(n_i, \theta) \frac{dpcm(n_i, \theta)}{d\theta} \tag{2.8}$$

If all firms are symmetric ($n_i = n$ for all $i \in I$) and no firm exits, market shares are not affected by a higher θ, but pcm will be lower for each firm: hence $dPCM/d\theta < 0$. Further, if competition is intensified by an increase in the number of firms (due to a fall in entry barriers), firms' market shares fall and again we find $dPCM/d\theta < 0$. Now consider an increase in the variance of efficiency levels n_i, that is firms become asymmetric. If θ is increased in the case with a positive variance in n_i, we get the following *reallocation effect.* As competition is intensified, market share is reallocated from inefficient to efficient firms. That is, s_i increases for firms with a high pcm_i and falls for firms with a low pcm_i. This raises the weighted average PCM if the first term in (2.8) outweighs the second (negative) term. This is most likely to happen when the variance in n_i is high and thus when concentration is high in the industry. It is straightforward to find theoretical examples where this indeed happens.In this chapter, we also use the Herfindahl index (H) as a measure of concentration. H is defined as

$$H = \sum_{i \in I} s(n_i, \theta)^2 = 1/N + NV \tag{2.9}$$

where N is number of firms and

$$V = \frac{\sum_{i \in I} [s(n_i, \theta)^2 - 1/N]}{N}$$

H is $1/N$ when all the firms have equal market shares (*i.e.* $s_i = 1/N$). Hence H depends on the number of firms. It is also straightforward to verify that a rise in the variance of n_i (for given average efficiency) leads to higher V and, consequently, higher H. Hence, as more aggressive interaction results in more variance in efficiency level this implies higher H. Implicating more concentration in a market, as more efficient firms increase their market share at the expense of less efficient firms. Consequently, a higher H can also go hand in hand with more competition.

30.[26] We calculate the Cournot equilibrium. The small dots in the right part of figure 2.1 give the relation between $\ln(\pi_i)$ and $\ln\left(\frac{c_i}{p_i}\right)$ before the increase in competition and the large dots the relation after competition has become more intense. After the increase in competition, the relation becomes steeper. Doing a simple OLS-estimation of PE with the data in this graph yields PE= 6.78 before and PE= 7.50 after competition is intensified. The number of active firms before and after equals 101 and 74 resp. Hence under the more competitive regime, inefficient firms can no longer enter and concentration increases. H incorrectly suggests that competition has become less intense, since the value increases from 0.010 to 0.014. The PCM falls here from 0.22 to 0.21. Hence, PE and PCM correctly indicate that competition has increased after d goes up.

The reason why H incorrectly suggests a fall in competition when the interaction between firms has become more aggressive is the *reallocation effect* and *selection effect*. With reallocation effect, we mean the reallocation of market shares between firms. As competition becomes more intense, market shares are reallocated from inefficient firms (with low initial market shares) to efficient firms (with relatively high initial market shares). Some inefficient firms may even go bankrupt due to the intensified competition and leave the market (*i.e.* selection effect). This raises concentration in the market incorrectly suggesting a fall in competition.

2.3.2 Simulations of competition measures

The examples in the previous subsection show that the three measures can diverge in response to more aggressive conduct. Therefore the simulations below entirely focus on this way to increase competition. Moreover, we examine only PCM and PE as measures of competition, since H is always wrong in this case. The impression of the preceding examples is that PE and PCM coincide in predicting the change in competition in both ways of intensifying competition. However, this is not always true as we will show below.

We use simulations to show that PE and PCM can point in opposite directions after an increase in competition in the case of changes in conduct. In such cases, PE usually points in

[26] This is a fairly standard way in which competition is parameterized in the literature. See, for instance, Aghion et al. (2005), Blanchard and Giavazzi (2003) and Vives (2008). The intuition is that product differentiation gives firms some market power. Since products are different, there is no head-to-head competition between firms. Making goods closer substitutes, reduces this market power and intensifies competition.

the correct direction. Further, we show that two variables (*i.e.* H and the reallocation effect) have some power in predicting when PCM incorrectly points to less intense competition. The simulations are based on the Cournot model with linear demand described above,[27] where $a = 40, b = 30$ and d equals (in the original situation) either 15 or 20. As above, competition is made more intense by increasing d with 10 (to 25 and 30, resp.). Firm i produces with constant marginal costs equal to c_i and faces a fixed cost that varies from $f = 0.004$ to 0.012. We assume that c_i is drawn from a lognormal distribution with mean 0.7 and the standard deviation (*i.e.* $st.dev.$) varies from 0.08 to 0.32.

For each combination of parameters we draw 110 values for c_i, as above. We calculate which of these 110 firms can profitably enter (pay the fixed cost f) under Cournot competition, where firms are assumed to enter in order of efficiency (most efficient firms first). Then we increase d with 10. This makes goods closer substitutes and is seen as an increase in competition. We derive the new Cournot outcome, again calculate PE and PCM. This we do 100 times for each parameter constellation (with each iteration we draw 110 new values from the cost distribution). We count the fraction of times that a measure gets it right. That is, after the increase in d competition has increased and PCM should decrease and PE should increase to signal this. The results are reported in table 2.1.

We use figure 2.2 to summarize the findings of the simulations. Each point in the two graphs is the result of 100 iterations for one particular choice of parameters.

The left part of figure 2.2 shows the fraction of these 100 cases in which PE and PCM correctly indicate an increase in competition (the parameter d is raised by 10) as a function of the average (over the 100 iterations) H before competition is intensified. A number of points follow from this figure.

First, PE performs very well with scores above 90% but it is not a perfect measure of competition. The estimated PE may fall in response to a rise in competition if the relation between “Log” profits and “Log” costs is non-linear. Then the first order Taylor approximation is no longer accurate. Entry or exit by firms relatively far removed from the other firms in the sample can then have a disproportionate effect.[28]

[27] These results do not only hold for the Cournot model, but also across other models as shown in Boone (2000a).

[28] In section 2.5.2 and chapter 3 we do a robustness check with respect to this non-linearity problem.

Table 2.1 Simulations results

d	f	$st.dev.$	PCM-score[a]	PE-score[b]	H[c]	Reallocation[d]	PCM [e]	PE[f]
15	0.004	0.08	1.00	1.00	0.0093	0.047	0.26	– 7.37
15	0.004	0.16	1.00	1.00	0.0102	0.070	0.28	– 6.87
15	0.004	0.24	0.83	1.00	0.0114	0.085	0.32	– 6.16
15	0.004	0.32	0.67	0.94	0.0124	0.095	0.36	– 5.63
15	0.008	0.08	1.00	1.00	0.0096	0.079	0.26	– 6.89
15	0.008	0.16	0.78	1.00	0.0110	0.094	0.30	– 6.02
15	0.008	0.24	0.45	0.97	0.0122	0.104	0.34	– 5.40
15	0.008	0.32	0.31	0.94	0.0132	0.115	0.37	– 4.95
15	0.012	0.08	0.82	1.00	0.0104	0.098	0.28	– 6.26
15	0.012	0.16	0.46	1.00	0.0118	0.108	0.32	– 5.47
15	0.012	0.24	0.23	1.00	0.0131	0.119	0.35	– 4.92
15	0.012	0.32	0.14	0.99	0.0140	0.128	0.39	– 4.53
20	0.004	0.08	0.91	1.00	0.0101	0.054	0.22	– 8.68
20	0.004	0.16	0.37	0.96	0.0121	0.069	0.27	– 7.34
20	0.004	0.24	0.09	0.93	0.0137	0.081	0.31	– 6.48
20	0.004	0.32	0.09	0.92	0.0150	0.091	0.35	– 5.87
20	0.008	0.08	0.30	1.00	0.0116	0.074	0.25	– 7.36
20	0.008	0.16	0.10	0.99	0.0135	0.086	0.29	– 6.28
20	0.008	0.24	0.06	0.97	0.0151	0.096	0.33	– 5.61
20	0.008	0.32	0.07	0.96	0.0164	0.103	0.37	– 5.11
20	0.012	0.08	0.06	1.00	0.0131	0.086	0.28	– 6.53
20	0.012	0.16	0.04	1.00	0.0150	0.096	0.32	– 5.67
20	0.012	0.24	0.00	1.00	0.0165	0.107	0.36	– 5.06
20	0.012	0.32	0.02	0.98	0.0177	0.116	0.39	– 4.64

[a] Fraction of cases in which PCM decreases (correctly pointing at increase in competition)

[b] Fraction of cases in which PE increases (correctly pointing at increase in competition)

[c] Average value of Herfindahl index before increase in competition

[d] Average value of reallocation effect

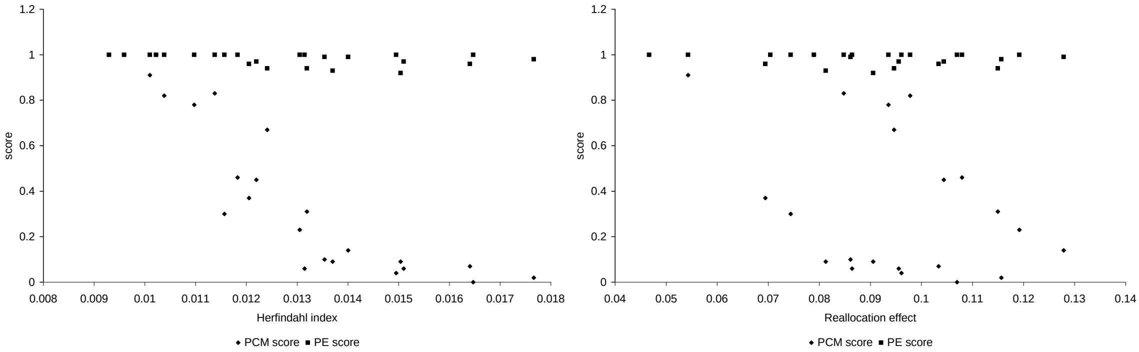
[e] Average value of PCM before increase in competition

[f] Average value of PE before increase in competition

Second, the PE score is not affected by the level of H, but PCM can point in the wrong direction for some parameter values with scores of below even 10%. Moreover, the higher H initially is, the more likely it is that PCM increases after an increase in competition. The reason is that high concentration is a necessary condition for a big reallocation effect. Intuitively, if there are 1000 small firms in a market, an increase in competition will not create much of a reallocation effect.

The right part of figure 2.2 relates the PCM and PE scores directly to the reallocation effect as defined in equation (2.10). The PE score is not correlated with the size of the reallocation effect, but the PCM score clearly decreases with this effect. A higher reallocation effect increases the probability that PCM goes up after an increase in competition.

Figure 2.2 Fraction of cases in which PE/PCM correctly indicate increase competition as function of average H (=left) or reallocation effect(=right)



In order to identify the reallocation effect in our firm level data, we have to decompose the change in the PCM for a market. Such aggregate change in PCM is made up from changes at the firm level with respect to PCM, but also from the consequences of the reallocation of market shares between firms in this market (see earlier textbox). Looking at the productivity literature, several methods have been developed to decompose an aggregate change (see Balk (2001) for an overview). We opt for a Laspeyres-type of decomposition. Therefore, we decompose the change in the PCM for a market in four different effects. In particular, we write

$$
\begin{aligned}
PCM_1 - PCM_0 = \sum_{i \in I_1} ms_{i1} pcm_{i1} - \sum_{i \in I_0} ms_{i0} pcm_{i0} = \sum_{i \in I} \{ & \underbrace{ms_{i0}(pcm_{i1} - pcm_{i0})}_{\text{within effect}} \\
& + \underbrace{pcm_{i0}(ms_{i1} - ms_{i0})}_{\text{reallocation effect}} + \underbrace{(pcm_{i1} - pcm_{i0})(ms_{i1} - ms_{i0})}_{\text{interaction effect}} \} \\
& + \underbrace{\sum_{i \in I_1 \backslash I} ms_{i1} pcm_{i1} - \sum_{i \in I_0 \backslash I} ms_{i0} pcm_{i0}}_{\text{entry and exit effect}}
\end{aligned}
\qquad (2.10)
$$

where $I_0(I_1)$ is the set of active firms before (after) the change in competition, $I = I_0 \bigcap I_1$ and $i \in I_1 \backslash I$ if both $i \in I_1$ and $i \notin I$. In words, the set I contains all firms that are active both before and after the increase in competition. Working with a balanced panel implies limiting the data to this set I. The set $I_0 \backslash I$ ($I_1 \backslash I$) contains firms that are active before the increase in competition but which are forced to exit after competition intensifies (firms that are active after the increase in competition but were not present before).[29]

In the simulations it turns out that PE and the within effect of the (aggregate or industry) PCM are strongly correlated. Since the within effect of PCM (by construction) is not affected by the reallocation effect, in principle it is a better measure of competition than the industry PCM. Hence, there are two other possibilities in which PCM can be (partly) corrected for this effect if firm level data are available.

First, the reallocation effect can be partly eliminated by using the unweighted PCM as measure of competition (as in Aghion et al. (2005)). This reduces the problem caused by the reallocation effect to a certain extent (as shown in Boone et al. (2005)) but does not remove it completely: an increase in competition tends to remove inefficient firms from the market with low PCM which raises the average PCM in the market. A disadvantage of the unweighted PCM is that the PCM of small firms gets a disproportionate effect on the industry (market) aggregate PCM.

A second solution to cope with the reallocation effect is to keep the weights s_i fixed at their baseline values. In other words, one can use the within effect as a measure of competition. We do not take this approach for two reasons. First, the within effect has to be based on a balanced panel (the set I in equation (2.10)). That is, if one wants to measure competition using the within effect consistently over a period of, say, 10 years one can only use data on the firms that are in the panel for all 10 periods. This limits the number of observations considerably if a data set is based on a (rotating) sample such as ours.[30] Second, in our data the within effect is a magnitude 10 smaller than the entry and exit effects (due to the fact that

[29] In the simulations, this set is empty, whereas in the real data, however, this set is not empty.

[30] Alternatively, one can calculate the within effect for consecutive years from t to $t+1$ and then with a new sample from $t+1$ to $t+2$ etc. In this way, fewer observations are lost. The disadvantage of this approach is that the reallocation effect plays a role again in the comparison of competition between t and $t+2$ as the base changes between those years. In this way, the within effect is not a consistent measure over the whole period. As explained below, PE is not affected by an unbalanced panel data set.

we use an unbalanced panel).[31] Hence due to the noise in the other effects, we cannot use the within effect in the data to benchmark PCM.

2.4 Data on competition measures

In this section we take a first look at the three measures PE, PCM and H in the real data. The latter based on Dutch firm level data from about 250 markets over the period 1993-2002.[32] We define a market to be a 3-digit SIC-code divided into small and medium sized firms (SMEs which have less than 50 employees) and big enterprizes (BEs which have 50 employees and more).[33]

Available data set

The estimates for PE, PCM and H are based on firm level data for the Netherlands. These data are derived from the annual survey for the Production Statistics (PS) by Statistics Netherlands. The survey gives complete coverage of firms with at least 20 employees, while firms with fewer than 20 employees are sampled. This chapter focuses on the period 1993-2001 (and 2002 for service industries). The data set has been constructed after matching the detailed accounting data over time. We have no data at our disposal on the agriculture and fishing industry, banking and insurance, public utilities and health care industries but otherwise we cover all industries in the Netherlands.

It turns out that the matched data set was not complete for all industries in manufacturing and wholesale trade. For some industries at the 3 digit SIC code, observations for certain years were missing for firms with size less than 100 employees. Therefore, we excluded all observations of these industries.

Unprocessed firm level data may contain errors for various reasons. In order to obtain reliable firm level data we performed several 'cleaning' activities (largely similar to Creusen et al. (2006a), see also chapter 3 for further details).

In our data set we do not have information on either quantity or price separately. Hence we cannot calculate marginal cost. Therefore we divide variable costs by revenue per firm

[31] As shown in table 2.3, in our data the average (standard deviation) of the within effect equals 0.02 (0.45), of the reallocation effect 0.02 (0.19), interaction effect 0.01 (0.11), the entry part of the change in active firms effect 0.33 (3.53) and the exit part 0.25 (0.21) where all effects are normalized by PCM.

[32] In chapter 3, we extensively explain how we estimate PE. The calculation of PCM and H in the data is straightforward and has already been discussed in section 2.3.

[33] We divide a 3-digit SIC-code into SMEs and BEs for two reasons: (i) policy interest for this distinction, (ii) it can be argued that for many industries, BEs operate on international markets while SMEs serve regional or local markets. T-test scores for PE suggest that there is a significant difference in degree of competition between SMEs and BEs, being larger for the latter (see below).

and per year assuming that marginal costs are equal to the average variable costs per unit of output. The theoretical model discussed in section 2.3 suggests that PE can indeed be estimated with this approximation of marginal costs. Moreover, as suggested by figure 2.1, to estimate the relation between profits and costs, we need not have the data on all firms in the market. Clearly, more data is always better, but we can still estimate the relationship reliably when we only have a sample of firms in the market. This is not the case for measures like concentration and PCM which only make sense if the whole population can be observed.

Table 2.2 gives the summary statistics for the three competition measures and some market characteristics that we use in the analysis hereafter. Here, we work with the full sample of markets.[34] Except for the import share, which is derived from the National Accounts of Statistics Netherlands, all indicators are based on the PS. Ideally, the number of observations in Table 2.2 should be 139*2*10=2780 (*i.e.* 139 3-digit SIC-industries divided into SMEs and BEs for the period 1993-2002). However, the full sample contains less observations: 2104 observations. This smaller set is due to that (i) for manufacturing industries data only runs to 2001; (ii) not for every SIC-code SMEs or BEs are available; (iii) some SIC-codes are absent in particular years.

Table 2.2 Overview of variables

Variable	Mean	Standard deviation	Minimum	Maximum	Observations
PE	7.03	5.21	– 5.47	39.07	2104
Δ PE	– 0.13	4.41	– 32.81	34.45	1851
PCM	0.18	0.10	0.02	0.85	2104
Δ PCM	0.00	0.05	– 0.50	0.61	1851
H	0.12	0.12	0.00	0.97	2104

We find that on average (over all markets and years) PE equals 7 (in absolute values): a one percent increase in costs leads to a seven percent reduction in profits. However, there is substantial variation in PE. In one market, a one percent increase in a firm's costs leads to a 39% fall in its profits. The average values for PCM and H equal 0.18 and 0.12, respectively. Moreover, the standard deviations of both PCM and H are much smaller than the one for PE. The variables ΔPE and ΔPCM denote first differences in PE and PCM. It is interesting to

[34] Chapter 3 also considers subsamples and other definitions for estimating PE.

Table 2.3 Decomposition ΔPCM using equation (2.10)[a]

Effects	Mean	Standard deviation	Minimum	Maximum
Within	0.018	0.447	– 0.719	16.715
Reallocation	0.017	0.187	– 0.955	1.797
Interaction	0.010	0.112	– 0.615	3.009
Entry	0.334	3.526	0.000	149.728
Exit	0.252	0.213	0.000	1.000

[a] The decomposition is based on 1814 observations. Each effect is normalized by PCM in the base (0) year.

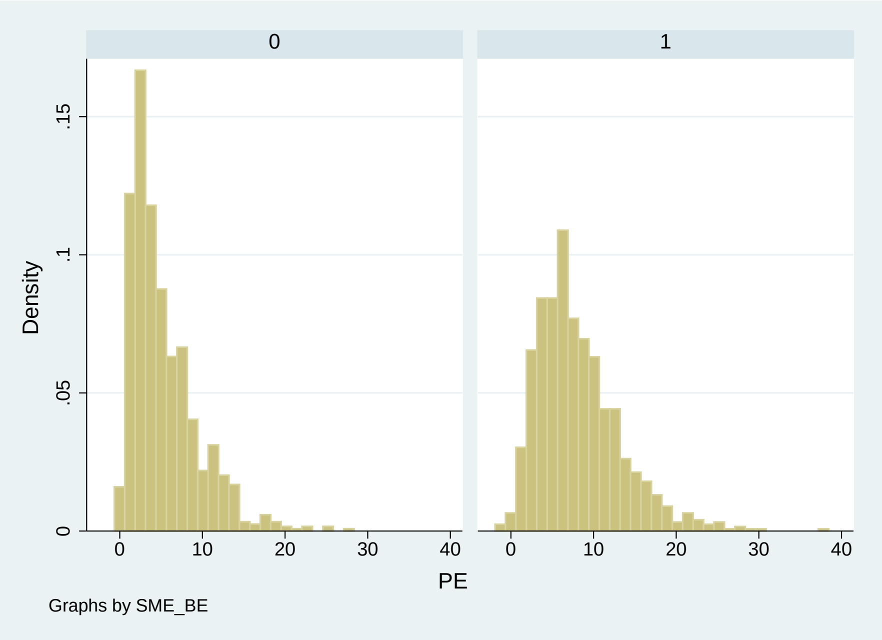
note that both variables are on average nearly zero. Table 2.3 gives the decomposition of ΔPCM using equation (2.10). Of this decomposition we will later use the reallocation effect to identify cases where changes over time in PCM and PE contradict each other. Due to our panel data structure, the entry and exit effects seems to be important for PCM. This corresponds to findings for decompositions of labor productivity (see Baily et al. (1992); Haltiwanger (1997); Bartelsman et al. (2003) and Baldwin and Gu (2006)).

Figure 2.3 summarizes the PE's that we find in the data with histograms. We give separate histograms for the two sub-markets: SMEs and BEs. As one can see in this figure BEs have substantially higher values for PE than SMEs (which is the main motivation for us to subdivide markets in this way). This is in contrast to the idea in policy circles that entrepreneurs and SMEs are key to economic performance. These firms supposedly increase productivity and competitiveness. Moreover, with respect to innovative change, they are believed to play an important dynamic role. In other words, these firms are claimed to face very intense competition and therefore have a big incentive to reduce costs and innovate. We find exactly the opposite. It is the big firms that face the higher values for PE. If their costs go up by 1% the percentage fall in profits is bigger. Note that this is not just a trivial size effect as we consider the percentage change in profits.[35]

Figure 2.4 gives the histograms for PCM. PCM tends to be lower for BEs than for SMEs, again showing that BEs are active on a more competitive market. Our interpretation is that in many markets BEs compete on a national market (or even international markets) while SMEs compete on local markets.

[35] It is obviously the case that the absolute change in profits due to an increase in marginal costs is bigger for a firm with a higher output level.

Figure 2.3 Distribution of PE in the Dutch economy (left=SME, right=BE)



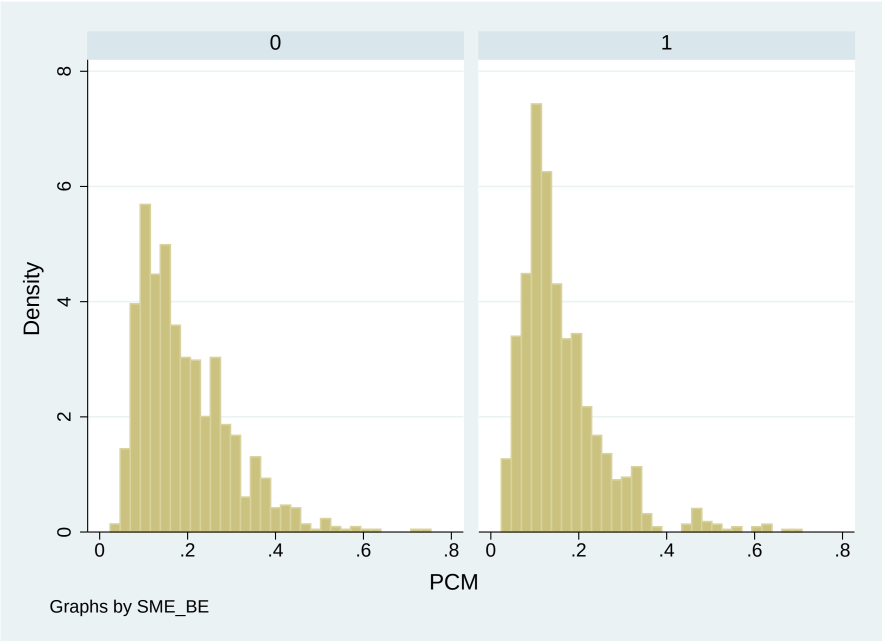
The histograms for H in figure 2.5 do not confirm the results seen for PE and PCM. The market for BEs tends to be more concentrated than the market for SMEs. Given that H is based on market shares, it is not surprising that BEs tend to be active on highly concentrated submarkets.[36] However, section 2.3 and the histograms above for PE and PCM clearly indicate that higher concentration should not (always) be associated with less intense competition.

2.5 Comparing measures of competition

This section first considers the cross section correlations between the three measures. As in the simulations and as suggested by figure 2.3 and 2.4, we find that PCM and PE are closely (negatively) correlated while H seems the "odd one out". Although this might suggest that PE and PCM always point in the same direction, this is not the case as we will show. We analyze the changes in PCM and PE over time and find that PCM tends to be different in

[36] Remember that market shares –and thus concentration– are calculated for submarkets consisting of a 3 digit SIC code and size class.

Figure 2.4 Distribution of PCM in the Dutch economy (left=SME, right=BE)



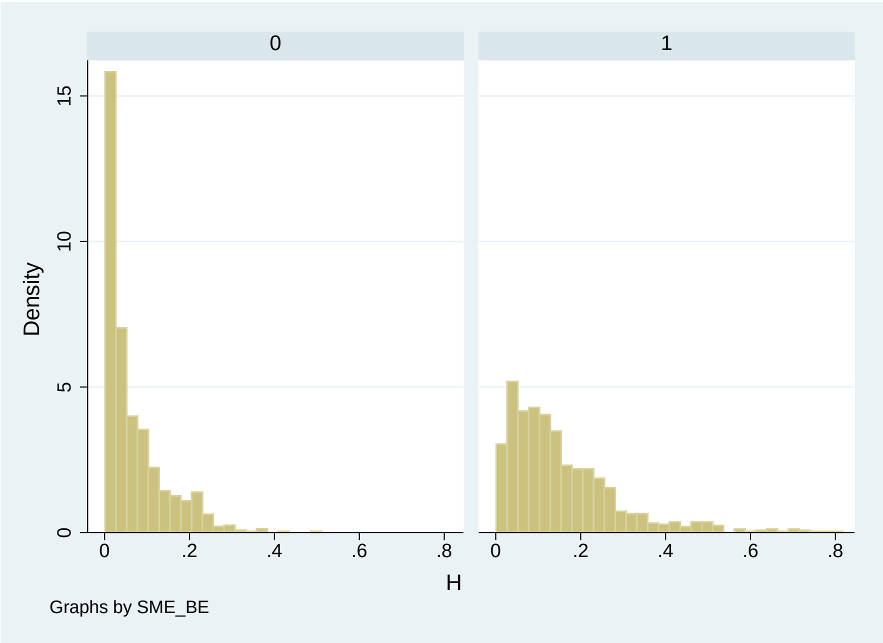
concentrated markets in accordance with the simulations.

2.5.1 Properties of competition measures

It turns out that the (direct) correlation between PE and PCM is negative and significant as one would expected. However, this could be a spurious correlation in the following sense. It could be that due to some market characteristics one measure is low and the other is high (and the other way around). If so, a negative correlation between the two measures may have to do with differences in these market characteristics rather than with agreement between the two measures on the underlying intensity of competition. To deal with spurious correlation we compare all three measures of competition in two steps. First, we relate them all to market characteristics. Then, we investigate the partial correlation between various measures conditional on the market characteristics.

Thus, in the first step, we perform a number of regressions in which PE (and other endogenous variables; see below) in market k at time t is explained through a number of market

Figure 2.5 Distribution of H in the Dutch economy (left=SME, right=BE)



characteristics that are assumed to be exogenous to competition.

$$PE_{kt} = \gamma_0 + \gamma_t + x'_{kt}\gamma + \varepsilon_{kt} \tag{2.11}$$

where x is a vector of market characteristics, the $\gamma's$ are parameters – with γ_t being calendar year fixed effects[37] – and ε is an error term.

As market characteristics we use the labor share in value added, the import share, the type of industry (dummy variable for manufacturing (1) versus services (0)) and the average firm size (dummy variable for large firms). We view these market characteristics as exogenous.

Labor income share is defined as total wage costs over gross value added – gross output minus intermediate inputs. In other words, it is the share of labor in the surplus created by labor and capital. We interpret a high labor income share as a property of the market that there are low capital requirements to enter the market. In this sense, we view a high labor income share as indicating low entry costs. The import share denotes the fraction of output sold on the domestic market from firms outside the Netherlands.

[37] The calendar year fixed effects are included to take cyclical effects into account.

Table 2.4 Properties of competition measures[a]

	PE	PCM	H	Labor prod.	Variance AVC	Number of firms
	(1)	(2)	(3)	(4)	(5)	(6)
Lab. inc. share	1.39 (10.6)**	– 0.49 (14.3)**	– 0.17 (3.7)**	– 2.63 (5.6)**	– 0.29 (5.8)*	1.46 (3.2)**
Import share	0.07 (0.7)	– 0.02 (0.9)	0.11 (3.1)**	– 0.09 (0.5)	0.02 (1.6)	– 1.60 (4.1)**
Manufacturing	0.40 (8.4)**	– 0.07 (5.5)**	– 0.01 (0.6)	0.12 (1.2)	– 0.08 (7.1)**	– 1.34 (5.8)**
Big enterprizes	0.19 (5.7)**	– 0.02 (2.4)**	0.08 (6.4)**	0.15 (3.4)**	– 0.04 (5.2)**	– 1.47 (8.8)**
R^2	0.385	0.642	0.189	0.376	0.185	0.503

[a] Based on 2104 observations (253 markets); all estimates contain fixed effects for calender years; absolute t-statistics in parentheses – corrected for clustering of observations across markets; a ** (*) indicates a parameter estimate significantly different from zero at a 5% (10%) level.

Nonetheless, we acknowledge that both characteristics are not completely exogenous to the intensity of competition. To illustrate, the intensity of competition in the product market can affect labor unions' bargaining power. If firms have market power, they may be able to affect the wage rate and the labor income share. Further, domestic markets where firms hardly compete are particularly attractive for foreign firms to enter, leading to a high import share. These caveats should be kept in mind. However, we believe that both explanatory variables are also driven by exogenous variation. The market's technology determines how much capital is needed to produce thereby affecting the capital income share and its complement the labor income share. Also, some products are easier to import than others which affects the import share. Markets where foreign products are close substitutes of domestic firms' products will face a tougher competitive regime. It is this effect that we try to capture.

We estimate equation (2.11) for PE, PCM and H. In addition to the referred market characteristics we estimate this relation for variables that we believe are rather closely driven by the intensity of competition: labor productivity, variance in average variable costs and the total number of (domestic) firms in the market. One would expect that in a competitive market, labor productivity is high while the variance in costs and the number of firms are small as inefficient firms cannot survive under intense competition.

Table 2.4 shows the estimation results.[38] The labor income share has a positive effect on

[38] Note that PE is divided by 10 in these regressions.

PE (see column 1). A high labor income share indicates low capital costs and hence it is easier to enter the market. The import share has a positive but insignificant effect on PE. The dummy variable for manufacturing industries also has a positive and significant effect on PE. Conditional on the other market characteristics there is more competition in manufacturing industries than in service industries, confirming what Creusen et al. (2006a) found for the Netherlands. Also in markets where large firms operate there is more competition.
The second column of table 2.4 shows the parameter estimates when PCM is the dependent variable. By and large the (significant) parameter estimates are very similar – though of course with opposite signs.

The third column of table 2.4 presents how H is affected by the market characteristics. As expected the labor income share has a negative effect on H. As less capital is required, it is easier to enter and concentration is lower. The import share has a positive effect on H. However, note that H is calculated on the basis of domestic revenues of domestic firms, the imports itself are not taken into account when calculating H. This may introduce a spurious, positive, correlation between import share and H. Such a positive correlation is indeed what we find. More imports on a market lead – *ceteris paribus* the size of the market – to less "space" for *domestic* firms. This tends to increase the domestic concentration. H is also large for markets with big enterprizes. Since PE and PCM suggest that markets with big enterprizes are more competitive, this suggests that more intense competition can go together with high concentration. More intense competition removes inefficient firms from the market thereby increasing H. We come back to this point below.

The fourth column of table 2.4 shows the results for labor productivity.[39] The average labor productivity is low in markets with a high labor income share. This may be due to the fact that such industries are labor intensive and therefore labor productivity is low. Labor productivity is high in markets with big firms and higher in manufacturing than in services. Import share does not have a significant effect on the average labor productivity. The fifth column shows that the variance of the AVC is influenced in the same (but opposite) way by market characteristics as the PCM.[40] This is not surprising as these measures are closely related.

[39] Labor productivity denotes the (unweighted) average labor productivity of the firms in an industry, where labor productivity is defined as gross value added per worker.

[40] Variance of AVC is the variance (over firms in a market) of the average variable costs per firm.

Finally, the last column in table 2.4 shows that the number of firms in the market is positively correlated with labor income share. This also suggests that a higher labor income share is associated with lower entry costs and hence more firms enter the market. The number of (domestic) firms is negatively correlated with import share, manufacturing and the market segment with big enterprizes. Since each of these variables are correlated with more intense competition (see columns for PE and PCM) this again indicates that more intense competition due to more aggressive conduct leads to fewer firms in the market.

Table 2.5 Partial correlation coefficients[a]

	PE	PCM	H	Labor productivity	Variance AVC	Number of firms
PE	–	– 0.147**	0.175**	0.091**	– 0.026	– 0.207**
PCM		–	– 0.007	0.154**	0.178**	0.096**
H			–	0.177**	0.008	– 0.571**
Labor productivity				–	0.101**	– 0.109**
Variance AVC					–	0.017
Number of firms						–

[a] The partial correlation coefficients are calculated holding the exogenous variables and the calendar year effects constant.

Table 2.5 shows the partial correlation coefficients between the three competition measures and other variables closely related to competition. As shown PE and PCM are not only negatively correlated through market characteristics. Also when keeping the market characteristics constant, there is a significant negative correlation between PE and PCM. This is mutually consistent, *i.e.* if one measure indicates more (less) competition so does the other. However, between PE and H there is a significant positive correlation. At first sight this seems inconsistent. After all, higher PE means more competition and a higher H means less competition. Yet, this confirms the idea introduced above: in a more competitive market, inefficient firms cannot survive and concentration goes up. Table 2.5 also shows that average labor productivity is positively correlated with PE. More intense competition weeds out inefficient firms and hence average productivity goes up after those firms have exit the market. Furthermore, PE is negatively correlated with the variance of the average variable costs and with the number of firms in the market. This also suggests that more intense competition weeds out inefficient firms thereby reducing the variance in costs.

PCM and H are negatively correlated, also suggesting that more intense competition in terms of lower PCM can go together with higher concentration. The partial correlations of PCM with variance AVC and number of firms are in line with the correlations of PE (with opposite sign). An interesting result is the positive correlation between PCM and labor productivity. This suggests that more efficient firms (higher productivity) have higher PCM for given mode of competition. Although PCM suggests that competition is less intense in more efficient markets, PE points to more intense competition in such markets. The hypothesis is that more intense competition affects the efficiency levels of firms leading to lower costs levels and higher PCM (given the price). Analyzing this point in depth is beyond the scope of this chapter. We leave it for future research to establish whether PCM can give the wrong impression in markets where firms can affect their cost levels.

The partial correlations of H with labor productivity and number of firms are consistent with the idea above that more intense competition removes inefficient firms from the market, thereby raising concentration and labor productivity while reducing the number of firms in the market.

As the average labor productivity is higher, ceteris paribus, the wider the range of AVC-levels that can be supported by a market. More firms in the market is correlated with lower average labor productivity. Finally, more firms on the market goes together with a higher variance in AVC.

2.5.2 When is PCM correct in measuring competition?

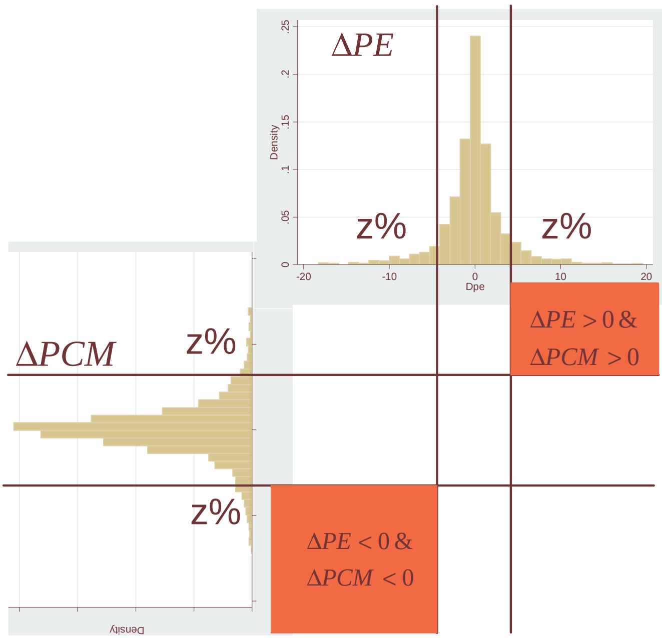
In the empirical analysis above we find that PE and PCM are affected by the same market characteristics and conditional on these market characteristics they are significantly negatively correlated. Nevertheless, in section 2.3 we have shown that there may be circumstances in which changes in PCM indicate, say, a fall in competition whereas PE shows an increase in competition intensity.

As shown in table 2.2, the average changes in both PE and PCM are close to zero. However, in particular markets there may be (big) changes which are not always mutually consistent in terms of changes over time in competition. It turns out that this happens in roughly one third of the cases.[41] To investigate this in more detail we first localize the markets in which

[41] Creusen et al. (2006a) for the Netherlands and Boone et al. (2005) for the UK also find such large number of cases.

there is an inconsistency between the two measures, *i.e.* markets where they are positively correlated from one period to the next. In these cases one measure indicates an increase in competition while the other measure indicates a decrease in competition intensity. We refer to these cases as being *strictly* inconsistent. However, if the changes in the measures are close to zero, the fact that they have similar signs does not matter that much. Such differences can be caused by observational errors and not by underlying changes in competition intensity. Only if both changes in the measures are substantially different from zero and with the same sign there is clearly something wrong. We focus on these cases in the following way.

Figure 2.6 ΔPE and ΔPCM are "very" inconsistent for market-year combinations in the grey areas



We define a dummy variable I_z which indicates whether or not ΔPE and ΔPCM are inconsistent, *i.e.* they have the same sign and are of sufficient magnitude. More specific we define $I_z = 1$, if

$$\Delta PE < \mu_{1,z} \& \Delta PCM < \mu_{2,z} \tag{2.12}$$

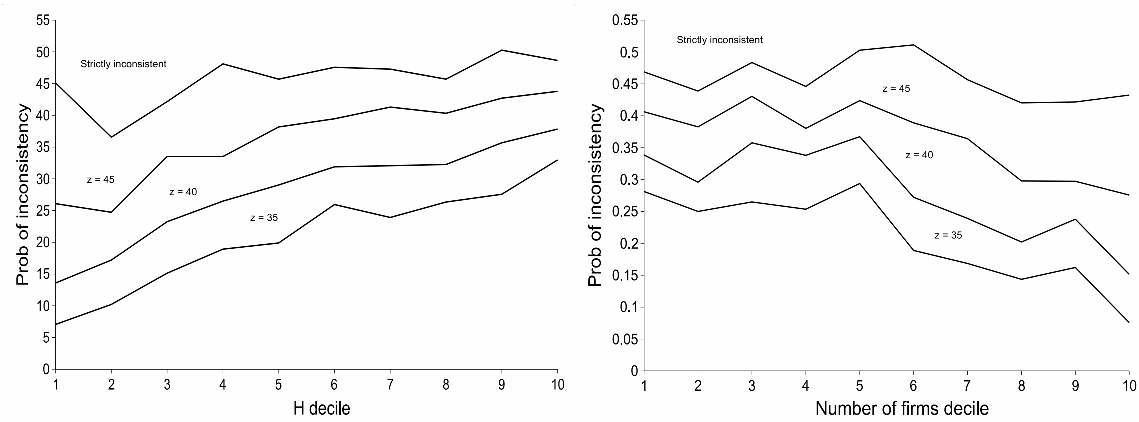
or

$$\Delta PE > \mu_{1,100-z} \& \Delta PCM > \mu_{2,100-z} \tag{2.13}$$

and $I_z = 0$ otherwise. Here $\mu_{1,z}$ is the value of the z^{th}-percentile of the distribution of ΔPE and $\mu_{2,z}$ is the value of the z^{th}-percentile of the distribution of ΔPCM. Hence $I_z = 1$ if both ΔPE and ΔPCM are either "strongly" negative or "strongly" positive. In these cases the two measures clearly contradict each other. This is illustrated by the shaded areas in figure 2.6. In addition, to investigate the importance of the reallocation effect in the change in PCM we define a new dummy variable "Big reallocation effect" which has a value of 1 if the reallocation effect (relative to PCM) is below the 25^{th}-percentile or above the 75^{th}- percentile of the distribution of reallocation effects.[42]

Now it is possible to investigate the determinants of the probability that the changes in the two measures are inconsistent, for various values of z. Figure 2.7 (left part) shows how the probability of inconsistency increases with H. This is hardly perceptible for cases with strict inconsistency ($z = 50$), but the increase is clear for low z values, *i.e.* when there is a big inconsistency. Similarly, the right part of figure 2.7 shows that this probability of inconsistency is decreasing in the number of firms.

Figure 2.7 Probability of inconsistency as function of deciles (left=the H-index; right= number of firms in market)



We investigate the determinants of inconsistency in more detail using a logit model to estimate the probability of inconsistency and relate this to H, number of firms and the realloca-

[42] As argued in footnote 31 the within effect is too small to use here to identify the reallocation effect. Since the reallocation effect is also quite noisy, we turn it into a dummy variable.

tion dummy. Table 2.6 presents the parameter estimates. Even if we consider the situation in which ΔPE and ΔPCM have the same sign – the changes are strictly inconsistent – H has a positive and significant effect (although this is not visible in the left part of figure 2.7). The number of firms in the market has a negative and significant effect on the probability of inconsistency for values of z below 45. Intuitively, with many small firms in the market, the reallocation effect will not be big enough to push PCM in the "wrong" direction (*i.e.* the first term in equation (2.8) does not outweigh the second term). Further, the effects of the reallocation dummy are significant for low values of z.[43] In markets with a high H, a low number of firms and a big reallocation effect we find that the probability of inconsistency between PCM and PE is large.

Table 2.6 Probability of inconsistency between ΔPE and ΔPCM; parameter estimates logit model [a]

	H-index	Big reallocation effect	Number of firms	% Inconsistent
Strictly inconsistent	0.60 (1.7)*	–	–	
	0.59 (1.6)	0.06 (0.8)	–	
	0.33 (0.7)	0.06 (0.7)	– 0.03 (0.9)	45.7
$z = 45$	1.52 (3.7)**	–	–	
	1.48 (3.6)**	0.16 (1.6)	–	
	0.70 (1.4)	0.15 (1.5)	– 0.08 (2.2)**	36.4
$z = 40$	2.17 (5.0)**	–	–	
	2.12 (5.0)**	0.25 (2.4)**	–	
	0.91 (1.8)*	0.23 (2.3)**	– 0.14 (3.3)**	27.9
$z = 35$	2.85 (6.3)**	–	–	
	2.79 (6.5)**	0.44 (3.9)**	–	
	1.53 (2.9)**	0.42 (3.7)**	– 0.15 (4.3)**	20.8

[a] Based on 1851 observations (250 markets); absolute t-statistics corrected for clustering of observations across markets; a ** (*) indicates a parameter estimate significantly different from zero at a 5% (10%) level. Time dummies are included.

We conclude that the reallocation effect is responsible for the inconsistency between the changes in PE and PCM. There is direct evidence because the probability of inconsistency

[43] We also investigated whether other variables used in tables 2.4 and 2.5 are important but none of them differed significantly from zero in any of the estimates.

increases with the size of the reallocation effect. There is also indirect evidence because the probability of inconsistency increases with H and falls with the number of firms. For this effect to be significant, we need to focus more on the tails of the distributions of ΔPE and ΔPCM ($z = 40$ and $z = 35$).

Robustness checks

To investigate the robustness of our estimation results, we run four alternative equations compared to our basic equation (2.1).[44]

The first alternative way to estimate PE is that we switch places for the dependent variable and the explanatory variable. In fact, this is one way to test the impact of measurement problems

$$\ln(c_{it}) = \alpha_i + \alpha_t - \tilde{\beta}_t \ln(\pi_{it}) + \varepsilon_{it} \tag{2.14}$$

In this case, PE is defined as $PE_t = 1/\tilde{\beta}_t$

The second alternative allows for a non-linear relationship between $\ln(\pi_i)$ and $\ln(c_i)$:

$$\ln(\pi_{it}) = \alpha_i + \alpha_t - \beta_{1t} \ln(c_{it}) + \beta_{2t}(\ln(c_{it}))^2 + \varepsilon_{it} \tag{2.15}$$

Due to this non-linearity, the results for the β's cannot be directly interpreted as a measure of market competition. Taking the first derivative of (2.15) with respect to c, we get

$$\frac{\partial \ln(\pi_{it})}{\partial \ln(c_{it})} = -\beta_{1t} + 2\beta_{2t} \ln(c_{it}) \tag{2.16}$$

which varies between firms in the market. A market value for PE can now be derived by using the market average of the marginal cost ($\bar{c}_{it}$) as follows: $PE_t = -\beta_{1t} + 2\beta_{2t} \ln(\bar{c}_{it})$.

The third alternative measure for PE is that we use a balanced panel instead of an unbalanced panel to make sure that our results are not driven by panel attrition. To be left with sufficient observations, we use a balanced panel for two subperiods: 1993-1997 and 1998-onwards respectively.

The fourth alternative is that we adjust our AVC concept accounting only for the labor costs and neglecting the costs for materials and other intermediate inputs. This relaxes the

[44] Chapter 3 elaborates more on alternative measures for PE.

problem of using the same variables to construct the left hand side and right hand side of equation (2.1).

Table 2.7 checks whether our main result is robust to these alternative specifications of PE. Indeed we find for all four alternatives that the probability of inconsistency is higher in more concentrated markets (higher H and lower number of firms). For alternatives 1, 3 and 4 it is also the case that a big reallocation effect raises the probability of inconsistency. In this sense, the main result in this section is robust to different ways in which PE can be estimated.

Table 2.7 Comparing alternative Profit elasticities – Probability of inconsistency between ΔPE and ΔPCM; parameter estimates logit model – $z = 35$[a]

		H-index	Big reallocation effect	Number of firms	% Inconsistent
Baseline		2.85 (6.3)**	–	–	
Alternative	1	3.49 (6.2)**	–	–	
Alternative	2	3.52 (6.8)**	–	–	
Alternative	3	2.56 (5.4)**	–	–	
Alternative	4	2.95 (6.3)**	–	–	
Baseline		2.79 (6.5)**	0.44 (3.9)**	–	
Alternative	1	3.45 (6.3)**	0.26 (2.2)**	–	
Alternative	2	3.49 (6.8)**	0.16 (1.4)	–	
Alternative	3	2.51 (5.4)**	0.28 (2.6)**	–	
Alternative	4	2.90 (6.3)**	0.26 (2.3)**	–	
Baseline		1.53 (2.9)**	0.42 (3.7)**	– 0.15 (3.4)**	20.8
Alternative	1	2.38 (2.8)**	0.24 (2.1)**	– 0.13 (2.6)**	20.4
Alternative	2	0.26 (0.5)	0.11 (1.0)	– 0.39 (8.3)**	25.2
Alternative	3	1.04 (1.7)*	0.25 (2.5)**	– 0.17 (3.5)**	19.7
Alternative	4	1.07 (2.0)**	0.23 (2.1)**	– 0.21 (4.5)**	23.8

[a] Based on 1851 observations (250 markets); absolute t-statistics corrected for clustering of observations across markets; a ** (*) indicates a parameter estimate significantly different from zero at a 5% (10%) level. Time dummies are included.

2.6 Conclusions

In this chapter, we introduce a new measure of competition: the profit elasticity (PE) – the percentage fall in profits due to a 1% increase in efficiency captured by (marginal) costs. An increase of this elasticity indicates an increase in competition because firms are punished more harshly (in terms of profits) for losing efficiency.

In general, PE and price cost margin (PCM) are often consistent as measures for changes in competition. This is certainly the case when competition intensifies through an increase in the number of firms in a market due to a fall in entry barriers, but also when competition becomes more intense due to more aggressive conduct by incumbent firms. In contrast, concentration rates like Herfindahl index (H) only correctly picks up the former one, whereas changes in firms' conduct are not correctly picked up due to a reallocation effect of markets shares between firms.

However, we argue that PE and PCM can be inconsistent as well if firms' conduct changes. Changes in PCM at the industry level are driven not only by changes in competition but also by market share reallocation among firms. To analyze this, we have compared their evolution over time for about 250 markets in the Netherlands. It turns out that PCM and PE point in different directions (one suggesting that competition went up while the other suggests that competition went down) in concentrated markets where the reallocation effect is important, *i.e.* when H is high and the number of firms is low. Simulations suggest that in such markets PCM can increase in response to more intense competition. Hence in highly concentrated markets *i.e.* in markets that are most relevant for competition policy and regulation, one should be careful using PCM as a measure of competition intensity.

Finally, we have found empirical support for the idea that more intense competition (due to more aggressive conduct by firms) removes inefficient firms from the market thereby increasing concentration. Such an increase in concentration should not be interpreted as a fall in intensity of competition. Further, more intense competition also tends to increase the average productivity of the remaining firms in the market (either due to a selection effect or because remaining firms are forced to invest more to reduce costs). This can also raise PCM while it is not a sign of weakened competition. Although the empirical evidence is based on Dutch data, we are confident that these results will be confirmed in firm level data sets in other countries.

All in all, our results have the following implications. Competition authorities and regulators should be careful to use observed increases in concentration and price cost margins in an industry as evidence of collusion or abuse of a dominant position. If the industry under consideration is already quite concentrated, such trends may actually be caused by rising intensity of competition. Estimating PE is then a straightforward way to determine whether this is the case or not. In this sense, just when H and PCM are needed the most they are not reliable. Hence, the PE is more robust from a theoretical and empirical perspective, having no bias in concentrated markets or markets whith substantial market share reallocation among firms. Further, because PE is such a robust measure of competition, we advocate its use in empirical work analyzing the effects of competition on, say, efficiency, innovation and unemployment.

3 Robustness of Profit Elasticity

3.1 Introduction

This chapter shows how to estimate the Profit Elasticity (PE) in practice. It also analyzes the robustness of this rather new competition indicator as described in chapter 2 and, for instance, in Boone et al. (2007a,b), and applied in empirical studies like the ones from Creusen et al. (2006a,b), and Van der Wiel et al. (2008)).

The intuition of this competition indicator is that an increase of PE indicates more intense competition because firms are punished more harshly in terms of profits for being inefficient. Firms with low values of average variable costs are more efficient than those with high values of average variable costs. In fact, assuming that the average variable costs capture the efficiency of a firm, the PE is based on the correlation between these costs and the firm's profits using a loglinear econometric specification and firm level data.

The aim of this chapter is to provide a guide for researchers how to measure PE in practice when firm level data is available and what issues one should be aware of. It discusses the statistical and econometric problems related to estimate PE. The chapter assesses what the effect on the estimated PE is under a range of conditions. These conditions include alternative model specifications, different econometric estimation techniques, and the impact of measurement errors and selectivity issues. When those conditions yield similar results, this indicates that PE is likely a robust measure for analyzing the developments of competition in practice. In that respect, the chapter raises two questions:

- How to estimate PE in practice?
- How robust are the results of PE?

Our theoretical framework provides no clear indication of the exact functional form of the relationship between efficiency and profits. In our preferred model, we assume a loglinear relationship between profits and average variable costs, where the slope of the parameter is the PE, our competition indicator, and this can be interpreted as an elasticity: the percentage increase in profits due to a 1% increase in costs.[45] A higher PE (in absolute terms) indicates more competition. The question arises whether this loglinear model specification is defendable.

Unlike the price cost margin (PCM) and the Herfindahl index (H), PE is based on an econometric specification estimated by using econometrics. We prefer to use a fixed effect model (FE) to estimate PE as we allow for the (unobserved) individual specific effects of firms being correlated with the explanatory variable. More precisely, given the firm specific effect, we want to know what the consequences of a higher or lower PE are for firms in a market. Yet, it is unclear whether this econometric technique of FE is more appropriate than other estimators. Here, we analyze the robustness of the estimation results of PE by comparing the results based on FE with the results of pooled OLS, First Differences and Random effect.

Similarly, we examine the robustness of our preferred model if we take into account measurement and selectivity issues related to our panel data. Researchers are dependent on the quality of the available micro data. In our case, we use firm level data from Statistics Netherlands to estimate PE. Although, Statistics Netherlands collects firm level data with care, still measurement errors in our key variables may bias the regression results of our competition indicator. As PE is based on panel data taken from a survey other panel data problems, like selection bias, might distort the results for PE as well. When particular firms are not present in the sample for reasons directly related to competition, this could create a bias in PE whenever the correlation between profits and efficiency levels of these (nonobserved) firms differ significantly from the firms in the sample were the regression is based on. Further, we assume that our efficiency measure is exogenous but that assumption can be challenged too. The cost variable could be endogenous, meaning that other (exogenous) variables determine this cost variable and this can generate a bias in the size of PE if our cost variable is correlated with the error term of our specification. This problem of endogeneity might be due to measurement errors, omitted or non-observed variables in our specification. Our proposed FE-model will not solve this problem if the unobservable variables are correlated with our cost variable and these variables are not constant over time.

[45] In fact we use a log-log model as both the dependent variable (*i.e.* profits) as well as the explanatory variable (*i.e.* average variable costs) are transformed to logs. This is also called a loglinear model as the relationship between both transformed variables is linear.

The main findings of this chapter is that the idea of the relationship between average variable costs and profits is to a large extent robust to different ways in which PE can be estimated. Our preferred model is neither rejected with respect to its functional form nor to its estimator technique. Hence, PE can be estimated at the industry level by using a loglinear model based on an FE model. We find statistically significant correlations between the results for PE based on FE and OLS, First Differences and Random Effect respectively. Moreover, the Hausman test (Hausman (1978) whether a fixed or random effects model is more appropriate, indicates that the firm specific effects should be seen as fixed rather than random. Further, the estimation results of PE are largely robust with regard to measurement errors. For instance, other definitions for profits and/or our efficiency concept significantly correlate with the results of our preferred model. Also, the results for PE are seemingly not biased because of the endogeneity problem. Finally, the effect of selectivity issues appears to be limited.

This chapter is structured as follows. Section 3.2 introduces our preferred model and the data we use. It also outlines statistical and econometric problems that will be discussed and assessed in next sections. Section 3.3 analyzes the importance of using an FE model in estimating PE compared to other estimation procedures. Next, we apply several sensitivity tests to check the robustness of our baseline specification (see section 3.4). Section 3.5 summarizes the main conclusions of this chapter.

3.2 Model and data

3.2.1 Introduction

This section presents very briefly the theory underlying PE and describes our preferred model to estimate PE.[46] It shows why this model opts for an FE model using firm level data. Then we turn to empirics and we start with a brief discussion of our panel data set, its limitations and we present some descriptive statistics. Finally, we discuss limitations of PE in practice if one wants to estimate this competition measure.

[46] More details can be found in chapter 2 and in, for instance, Boone et al. (2010a).

3.2.2 Preferred model for PE

We are interested in the relationship between firms' profits and (marginal) costs that capture firms' efficiency in a market. A priori, we have no idea about the functional form of this relationship. We have a preference for the loglinear form since this form is convenient in terms of economic interpretation. We have the following specification in mind

$$\ln(\pi_i) = \alpha - \beta \ln(c_i)$$

Where π denotes profits, c is costs, and i a firm. With this specification between $\ln(\pi_i)$ and $\ln(c_i)$, the slope β can be interpreted as an elasticity and is our measure of competition. Moreover, this loglinear model imposes constant elasticities for all firms in a market.[47] In our case, it indicates the percentage fall in profits due to a one percent increase in costs. Implicitly, we assume that the effect of fiercer competition on the *level* of profits differs among firms.

The econometric specification to start with could look like

$$\ln(\pi_{it}) = \alpha - \beta \ln(c_{it}) + u_{it} \tag{3.1}$$

We argue that it is not plausible to assume that u_{it} is *IID* (Independently Identically Distributed) random error over time as assumed in OLS.[48] Firms are very heterogenous (see *e.g.*, Bartelsman and Doms (2000); Van der Wiel and Van Leeuwen (2003); Bartelsman et al. (2004) and Van Leeuwen (2009)). We state that it is important to deal with heterogeneity across firms since firms differ widely in efficiency performance. Just like humans, firms vary in many aspects, amongst others: applied technologies, labor composition, R&D - and innovative efforts, age (*i.e.* young versus old firms), and scale and scope (*i.e.* small versus large product assortment). The reasons for these differences might be: first mover effect, talent, incentives or simply luck. They lead to higher profits for given c_i.

We prefer to use an FE-model to estimate PE because this type of econometric model has a number of advantages. The most important one is that it solves the omitted or unobserved variable problem that could be relevant. This non observed heterogeneity may also drive differences in efficiency levels across firms.

[47] In section 3.3, we test whether this loglinear specification makes sense.

[48] *i.e.* E[u_i]=0, var(u_i)=σ^2 and Cov (u_i,u_j)=0.

It is, therefore, more appropriate to assume that the error term in equation (3.1) may include unobserved things like firm specific issues (α_i), time-specific effects (T_t) and an idiosyncratic error (ε_{it})

$$u_{it} = \alpha_i + T_t + \varepsilon_{it} \tag{3.2}$$

We include those firm and time fixed effects for the following reasons.

The firm specific effect controls for non-observable variables that may have an impact on the relationship between profits and costs. We assume that the unobserved individual effects α_i can be correlated with costs c_i: α_i is not necessarily zero given c_i (= $cov(c_i, \alpha_i) \neq 0$). Moreover, we assume that costs are strictly exogenous (conditional on the unobserved individual firm effect). Stated otherwise, all c_{it} are independent of all ε_{it}.

$$\mathrm{E}\left(\alpha_i \left| c_i \right.\right) \neq 0$$

and

$$\mathrm{E}\left(\varepsilon_{it} \left| c_{i1}\right., c_{i2}, \ldots c_{iT}, \alpha_i\right) = 0$$

Indeed, if those non-observable variables are correlated with the dependent variable then the results of PE are biased using pooled OLS method. For instance, management quality including experience can be such factor but is most times hard to observe or missing in surveys of Statistical Offices. Nevertheless, differences in management skills can lead to differences in performances in terms of profits that are not (directly) related to the extent of competition and not reflected in the variable that reflects efficiency (see *e.g.*, Siebert and Zubanov (2008)). For instance, although the efficiency level in terms of cost per unit of output can be identical, CEO's with larger social networks or firms with better marketing departments than their counterparts may realize higher profits. If not taken into account, this could bias the estimation of PE.

Moreover, including firm specific effects, we allow for the fact that we cannot perfectly observe the relevant values for firms' profits and costs. For instance, a firm may produce other products than the products for the market under consideration. Statistical Offices (or another agency gathering the data) may decide to classify these sales and costs of other products under the same heading (*i.e.* industry classification). This may create a bias that can be controlled for with a firm fixed effect.

Taking account of a time-fixed effect in our preferred model is related to the following issue. The time fixed effect captures the effect of all variables that do not vary over the individual firm but vary over time for each firm in the sample. It corrects, for instance, for inflation, business cycle effects or other exogenous shocks that may have an effect on the profits of each firm but not on our cost measure as we will see later on.[49]

To study the determinants of product market competition or investigate the effects of competition on innovation and productivity, one needs a competition measure that varies both over time and across industries (see chapter 5 and chapter 4 for an application). Therefore, using the fixed effect estimation technique and including a time interaction effect with β, we estimate for each market an equation of the form

$$\ln(\pi_{it}) = -D_t\beta_t \ln(c_{it}) + \alpha_i + D_t + \varepsilon_{it} \tag{3.3}$$

where i denotes the firm, t the year of observation and D are year dummies.

The FE-model assumes the same slope of β for each firm i in year t with constant variance across firms within a market (*i.e.* ε_{it} is *IID*).

3.2.3 Panel data

To estimate PE, firm level data is needed. With micro data, this new competition measure needs exactly the same variables as PCM, i.e. data on revenues and variable costs. But one key advantage of the PE is that not all observations for each firm in a market are needed. In theory it is argued that an increase in competition raises PE for any three firms (see Boone (2008)). This property of PE is useful as it allows one to use data sets where not all firms in the industry are sampled. A sample of firms is enough to estimate the slope of equation (3.3) as a rise in competition raises the profits of any firm relative to any other firm that is less efficient. Hence, one can easily run the regression with an unbalanced panel, this in contrast to concentration measures like H which are difficult to apply and to interpret if not all firms in the industry are observed over time. Also, the PCM might be biassed in that case as well.

A panel data set with individual firm level data makes it then possible to analyze firm's

[49] To some extent, those time dummies also cope with autocorrelation in the error terms.

competitive behavior over time and across firms for a market. In that case, one can disentangle changes within firms over time and the cross-sectional information reflected in differences between firms. With respect to our model, there are two key data requirements for employing the FE-model. First, for each firm in the panel we must have at least two observations on the dependent and independent variable, otherwise this firm drops out of the analysis. Second, these observations must be different over time for having within variation.[50]

For this chapter, we use firm level data for the Netherlands. These panel data are derived from the annual survey for the *Production Statistics* (PS) by Statistics Netherlands. The survey gives complete coverage of firms with at least 20 employees, while firms with fewer than 20 employees are sampled. This chapter exploits data from the period 1993-2006. The data set has been constructed by Statistics Netherlands after matching the detailed accounting data over time. We have no data at our disposal on the agriculture and fishing industry, banking and insurance, public utilities and health care industries but otherwise we cover all industries in the Netherlands. Moreover, not every industry is present from 1993 onwards. Particularly, for services industries, most data start in 2000.[51]

As discussed in section 3.2.2, to measure PE, we need two variables per firm: its variable profits and its costs, the latter as indicator for the efficiency level of that particular firm. This efficiency should be one dimensional and firms should compete on a level playing field. As noted by Boone (2008), in case of two dimensional efficiency, an increase in competition forces firms to focus on its most productive activity. This can blur the relationship between profits and efficiency we want to estimate. We come back to this issue below.

In principle, we use the so called 'lean definitions of both variables as this definition is more in line with the former statement than the so called 'wide definition' as the latter also includes activities that are often not directly related to the core business of firms (see also textbox *Different definitions for profits and costs possible*). To capture the efficiency level per firm, we use the average variable costs (AVC) calculated as the variable costs over revenues as an approximation, because data on marginal costs is absent. Notice that we cannot measure

[50] If both the dependent and the independent variable display little variation over time, then the FE-model can produce relatively inefficient and therefore unreliable results (see Zhou (2001)).

[51] Note, that the sample we employ in this chapter differs from the one in chapter 2, as the former is based on the longitudinal database recently constructed by Statistics Netherlands. Moreover, chapter 2 only uses time series up to 2002, whereas this chapter employs data up to 2006. Hence, the results for PE are not directly comparable between the chapters.

the variable cost per unit of output as data on output prices are not available. In general, AVC are also decreasing in efficiency.[52]

Different definitions for profits and costs possible

The available data set allows us to construct both key variables (*i.e.* profits and efficiency) in different ways. Here, we distinguish between two types of definitions: lean and wide.

The so called *lean definition* approach tries to correct for activities that are not the core business of firms by leaving them out. In a sense, this definition tries to correct for issues that violate the assumption that efficiency should be a one dimensional variable. To be more specific, the lean definition of the variable profits is defined as: total revenues (=O000000) minus net turnover of other activities (= V21200H) minus the variable costs.[a] Here, the variable costs are calculated as operating expenses (=LH310000) minus depreciation costs (=F110000) minus the value of purchased commodities (=I110000). Put differently, this cost measure includes the sum of the labor costs and the intermediate inputs except the value of purchased commodities. Hence, it includes any costs that are seen as variable in the sense that those costs vary with small changes in production. More precisely, labor costs are defined as the salary of employees including social security charges and extra allowances. Intermediate inputs consist of costs of inputs like materials, energy and marketing that are related to the amount of output.

In case of the *wide definition* approach, the net turnover of other activities and the value of purchased commodities are included in the profits and variable costs respectively.

Variable	Wide definition	Lean definition
Total revenues	O000000	O000000- V21200H
Variable costs	LH310000 -F110000	LH310000 -(F110000+I110000)

[a] The code refers to the one used in the questionnaire of Statistics Netherlands.

Cleansing the data

Unprocessed firm level data may contain errors for various reasons. Those errors may include values incorrectly labeled, invalid values and duplication of data. Although Statistics Netherlands applies its own quality controls to overcome such processing errors, additional activities are still required. In order to obtain reliable firm level data we performed several 'cleansing' activities. We removed:[53]

- Observations of firms with no turnover and employment (736)

[52] As our AVC-measure is always lower than c for every firm, the PE will be upward biased.

[53] The number of observations that were unreliable and had to be removed are within brackets.

- The second observation of the same firm in one year (12)
- Observation of year t+1 if a firm has identical output and employment data (or value added) in two consecutive years (1442)
- Observation of firms with negative variable profits (59143)
- Observations of firms with negative intermediate inputs (14)
- Observations of firms with huge changes in key variables as output and employment. In particular, firms with more than 500% increase in turnover or employment or respectively decrease by more than 80% in these variables (2135)[54]
- Finally, due to confidentiality requirements of Statistics Netherlands, we had to remove 3-digit SIC industries if less than 5 firms per year were available (1691)

Particularly, the last three activities may lead to selectivity problems creating a bias in PE as the 'non-selected' firms may differ in their behavior from the selected firms. We come back to this issue in section 3.4.

Table 3.1 Comparison of uncleaned and cleaned data set, 1993-2006

Variable	Uncleaned data set	Cleaned data set
Number of observations	387575	322402
Number of firms	135561	121561
Number of branches	165	154
Average firm size sample	71	69
Number of workers (x1000)	37407	30996
Profits per firm (x1000)	258	292
Labor productivity (x 1000)	97	112
AVC	0.61	0.54

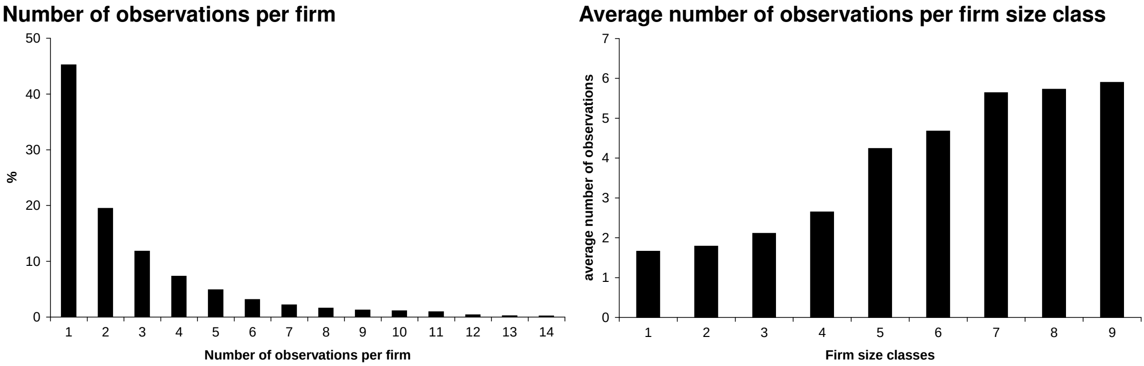
Note: Average firm size is the number of workers per firm; Labor productivity is defined as value added per employee; AVC is average variable costs.

Table 3.1 shows that the consequences of all those cleansing steps are limited. All in all, approximately 65,000 observations (i.e. approximately 17 percent) have been removed from the uncleaned data set to obtain a cleaned data set for further analysis. This cleaned data set contains more than 320,000 observations over the period 1993-2006 based on informa-

[54] Only observation $t+1$ is removed. This is to some extent arbitrary as one could also choose observation in year t.

tion of about 121,000 individual firms in the Netherlands from 154 industries at the 3-digit SIC-level.[55] Hence, on average we have at our disposal almost three observations per firm as required for applying an FE-model (see also figure 3.1). However, for most firms (approximately 45 percent of all firms), only one observation is available so they cannot be used in the FE-model because at least two observations are required. Particularly, firm size class 5 and lower (firms with less than 50 employees) have on average less than three observations per firm (see right part of figure 3.1).

Figure 3.1 Number of observations cleaned data set



The profits and labor productivity level per firm are higher in the cleaned data set, whereas the average variable costs are somewhat lower compared to the uncleaned data set. Profits being higher in the cleaned data set is what we would expect ignoring firms with negative profits. Higher labor productivity and lower AVC point to relatively more efficient firms in the cleaned data set suggesting that our cleansing steps could result in selectivity problems as we removed relatively more inefficient firms.

Applying the FE-estimation procedure[56] on estimating equation (3.3) at the 3-digit SIC-level, table 3.2 presents the key results for PE (in absolute value) based on three panels: uncleaned,

[55] If a firm change its production activity considerably, Statistics Netherlands reclassify this firm to a more appropriate SIC-code. To have a larger data set available for the fixed effect estimator, we arbitrary fixed the SIC-code to the highest in period 1993-1997 as well as period 1998-2006.

[56] Similar results could also be obtained by using the so called LSDV (=Least Square Dummy Variable)-estimator. This approach includes a dummy for each firm in the equation. As in our case N is very large the FE or within estimation procedure is more convenient than using dummies for each firm in the panel. Moreover, each dummy in

cleaned, and limited.[57] The limited sample includes the results where PE is not negative. For all three samples, PE is on average around 5, but there is a large variety across markets according to the size of the standard deviation. Particularly, the minimum value of PE can be very substantial.

Table 3.2 FE-model: overall results PE for different samples at 3-digit SIC-level, 1993-2006

	Observations	PE				T-value			
Panels		mean	st.dev	minimum	maximum	mean	st.dev	minimum	maximum
Uncleaned	2208	5.0	0.7	– 1561	94.3	6.7	0.6	– 6.3	43.6
Cleaned	1946	4.6	0.8	– 1170	60.1	7.3	0.7	– 6.2	43.6
Limited	1925	5.8	0.7	0	60.1	7.3	0.7	0	43.0

Note: Limited sample includes only positive PE-results.

The size of the average t-values suggests that the PE-results of our FE-model are very significant.[58] Getting rid of implausible negative results for PE (nearly 1 percent) at the 3-digit level, the overall picture does not change dramatically using the limited sample instead of the cleaned sample. Henceforth, we employ the cleaned data set for further analysis.

3.2.4 Ins and outs of PE in practice

The availability of panel data allows us to use an FE-model for the Dutch economy as we have more than one observation available (see figure 3.1) and the key variables profits and efficiency can be made operational. This section discusses some limitations of the PE if one wants to estimate this competition measure in practice.

The data quality of panel data may limit the application of PE in practice (*i.e.* "rubbish in rubbish out"). For instance, a priori we do not know how much of the difference in either efficiency levels or profits of firms is real and how much comes form errors in variables. Stated

LSDV removes one degree of freedom. If needed the individual effect of each firm can be computed from the FE approach.

[57] Notice that these results are based on aggregated (unweighted) results from the 3-digit level. For convenience, we show the PE results as positive figures.

[58] Note that leaving the firm specific dummies out of equation (3.3) the econometric statistics like t-values are not correctly measured due to wrong number of freedoms.

otherwise: are we picking up the right things in terms of development of competition? The following list of issues might bias the results of PE when not taken into account sufficiently:[59]

- Pure econometric problems
- Defining the relevant market
- Panel data problems

Each issue can have an unintended impact on the econometric estimates of PE and have to be verified in practice. In fact, if researchers want to test the validity of their hypothesis, those (econometric) steps are always needed, independently from estimating PE. On the other hand, econometric panel techniques provide the opportunity to consider issues like selectivity and measurement errors (see below).

Pure econometric problems

The econometric problems include the following:

- Unobservable factors including heterogeneity of firms and omitted variables
- Heteroskedasticity and serial correlation
- Choice of exogenous variable(s): Endogeneity problem
- Functional form of the model

We examine all those econometric issues to investigate whether our preferred FE-model for estimating PE is the best choice given the data set we have at our disposal. First, with regard to the unobservable factors, the question arises: do the firm specific and time specific effects really matter? In section 3.3, we show that these effects matter when compared to the results of three other econometric estimation procedures (*i.e.* OLS, First Difference, and Random Effect).

Second, we also take into account econometric problems related to heteroskedasticity and serial correlation. The former means that the assumption that the variances of the errors across individuals are identical no longer holds. Autocorrelation may occur in case the data set has a time dimension such as ours. It means that two or more consecutive error terms

[59] It should be emphasized that most of these issues also affect the measurement of the traditional competition indicators, except for the econometric problems. The latter is only relevant for PE as this indicator is based on an estimation of an econometric specification, while those other measures can be directly calculated from the data.

are correlated. Particularly, panel data observe the same firms over time and therefore these observations may be dependent. Both issues point to misspecification of the model in the sense of incorrect functional form or omitted variables and *IID* is harmed. In that respect, we also analyze whether the AVC as indicator for efficiency is endogenous and could bias our estimation of PE as it can be argued that this cost concept is not exogenous (see section 3.3.5). Finally, we test the functional form because it is not directly obvious that the loglinear functional is better than an alternative functional form given our data set (see section 3.3.1).

Defining the relevant market

Delineating the relevant market on which firms are competing is a notorious problem for each competition indicator, likewise for PE. The issue of defining the relevant market is particularly important for competition authorities detecting cartels and market abuse. This delineating of the relevant market requires information on the substitutability of the products and product attributes, and sometimes on the region in which the suppliers operate. However, in practice, this is often impossible to determine due to lack of appropriate detailed data as is the case in our situation.

In that respect, besides observing the efficiency level of a firm, two other assumptions are needed for PE to work correctly as competition indicator: (i) efficiency is one dimensional, (ii) firms compete on a level playing field (see Boone (2008)). These assumptions are also related to defining the relevant market and if these assumptions are violated, then the PE is not a perfect measure of competition.[60]

The first assumption is associated with the case where firms produce more than one good. The more aggregated the data become, say at the two digit level, the more likely it is that one firm is more efficient in producing one good and another firm more efficient in producing another good within this two digit category. In that case, efficiency might no longer be a one dimensional variable (see Boone (2008)) blurring the relationship with profits. Due to data availability, we have to assume that our efficiency measure indicates that the more efficient firm is more efficient in the production of all the goods.

The second assumption for measuring PE in the right way is related to the issue that some firms can face tougher competition than others in a market due to an unlevel playing field as

[60] It should be emphasized that the same applies to PCM.

more asymmetries than their costs levels might exist. This could be if firms face differences in taxes, but in our case we try to control for that as much as possible by using a similar cost definition including (Dutch) taxes.[61]

But still the data may give problems we cannot control for. For instance, entry by firms with efficiency levels far removed from other firms in the sample, so here we are really speaking of outliers, can have adverse effects on PE as referred to in chapter 2.[62] This is the case if this 'inefficient' entrant sells products with relatively high prices reflected in high profits. Then, the loglinear form for our relationship between profits and efficiency as assumed in equation 3.3 can be no longer accurate. As a result, the estimated β may fall or rise in response to a rise in competition if the relation between log profits and log costs is non-linear. But those exceptional entrants are not really the firms we want to examine with respect to competition.

Summarizing this discussion, if the data set allows then PE should be calculated at the level of aggregation that corresponds the best to the relevant market. We examine this issue in more detail in section 3.4.1.

Panel data problems

The limitations of panel data include: attrition problem including non response or low response rates, missing data and sample selection. The attrition problem is that one looses participants due to the fact that respondents, in our case firms, may abstain from response, go bankrupt or move to another market. Hence, the amount of nonresponse in panels increases in subsequent waves. These issues, particularly non response, could bias the results for PE if those nonobserved firms behave differently from those observed.

To some extent, Statistics Netherlands already tries to mitigate this attrition problem by applying the method of rotating panels. It refreshes the sample for firms smaller than 20 employees. This decreases the attrition bias as it is aimed to increase the response rate.

[61] Another example of an unlevel playing field is the difference between Dutch municipalities in allowing Sunday shop opening hours.

[62] Studies from, for example, Geroski (1995) and Santarelli and Vivarelli (2006) show that entry of new firms is heterogeneous with real (innovative) entrepreneurs being found together with market churning including 'revolving door' firms (*i.e.* over-optimist gamblers and escapees from unemployment).

However, it implicates that a complete history of observations per (small) firm is not available. Moreover, the number of firms in the survey sample has been significantly reduced over time due to reducing the administrative burden for firms. The attrition problem is partly related to the issue of selectivity problems in panel data. If either the nonresponse or our selection conditions are endogenously determined – meaning that they are related with the dependent variable –, this could create a bias in the estimation of PE. Note that missing data of firms pose no problem if they are randomly missing as they do not affect the slope between profits and costs. Further analysis of this issue is done in section 3.4.4.

A special issue in the limitations of panel data are *measurement errors* in the dependent and independent variable that could bias the results for PE, and hence the intensity of competition. Measurement errors may arise because of wrong responses due to unclear questionnaires or deliberate distortion of responses, or processing errors by Statistical Offices. Measurement problems also include the approximation of the efficiency concept. For instance, firms incorrectly fill in the questionnaire on depreciation costs (*e.g.* economic value versus tax value) or the labor costs do not include all costs including the payments in kind such as free lunches for their workers. Below in section 3.4.2, we explain that only random errors in the independent variable affect PE.

3.3 Estimating PE in practice

3.3.1 Testing for the functional form

As stated, our preferred model is loglinear assuming a (multiplicative) form with constant elasticities within a market. The simulations in chapter 2 show that PE based on a loglinear model performs well with scores of 90 percent or even higher predicting the right change in competition development.

A priori, however, the functional form is not settled and definitely not based on economic theory. The scarce empirical literature makes clear that the functional form of the relationship between profits and efficiency is not determined in practice yet. We refer to a number of examples. Boone et al. (2005) use a relative profit measure that deviates from the one proposed in this chapter. They relate the logarithm of a firm's profits with its costs without logs. So, they assume a non-linear relationship with different elasticities for different cost levels (*i.e.* different firms), while this chapter as well as, for instance, studies from Creusen et al. (2006a) and Van der Wiel et al. (2008) assume a linear relationship in logarithms with the same elasticity for firms in a market. Maliranta et al. (2007) employ both a linear and loglinear functional form to measure the competition intensity in Finland. Below, we put to a test whether the loglinear functional is better than an alternative functional form given our data set using three tests: (i) Ramsey Reset test, (ii) Box-Cox transformation test, (iii) piece-wise regression.

Ramsey Reset test: test for misspecification

The Ramsey specification error test (see Ramsey (1969); Patterson (2000) and Verbeek (2004)) or Ramsey Reset test is a general model (mis-)specification test that examines whether non-linear combinations of the estimated values for profits explain this exogenous variable. This test indicates that there may be some form of misspecification but it does not provide any indication of what the correct specification should look like.

Actually, the Ramsey Reset test is a test of a linear specification against a non-linear specification. It uses the predicted value of the dependent variable of the basic equation in a second equation. If the second and further exponents are significantly different from zero (by means of an F-test) than missing variables or wrong functional form could be an issue. The following regression is estimated

$$y = \beta x + \beta_1 \hat{y}^2 + \ldots + \beta_{k-1} \hat{y}^k + \varepsilon$$

The intuition behind the test is that, if non-linear combinations of the explanatory variables have any power in explaining the exogenous variable, then the model is misspecified.

If the null-hypothesis that all regression coefficients of the non-linear terms are zero is rejected, then the model suffers from misspecification of the functional form. Based on the F-statistic, we find no evidence for that. The results for the H0 (= model has no omitted variables) is as follows: F(36, 237461)=163.93; Prob>F = 0.0000. This means that the F-test statistics is not greater than the critical F-value. Hence, the null hypothesis that the correct specification is loglinear cannot be rejected. It suggests that the true specification of our model is loglinear.

Box-Cox transformation test

The Box-Cox transformation is another way to test the functional form of our relationship (see *e.g.*, Box and Cox (1964) and Verbeek (2004)). In fact, this method finds the maximum likelihood estimates of the parameters of the Box-Cox transform providing an indication for the functional form of both the dependent and independent variables.[63] The size of the fitted parameters for the dependent and independent variables, θ or λ, gives an indication for the functional form. For example, θ=0 means loglinear transformation, whereas θ=1 means linear or no transformation.[64] So, if the hypothesis of θ=0 is accepted, we should use log(profits) as the dependent variable. If, however, θ=1 is accepted than we should use no transformation.

As table 3.3 shows, the estimate of θ is – 0.034 when only the dependent variable is being transformed suggesting that a log transformation is the best option to proceed with for profits. For the explanatory variables, the estimate of λ is –0.5255 suggesting that either no transformation or log transformation can both be sensible. Transforming both sides, the estimate of λ is very close to zero pointing towards a loglinear model. Moreover, the likelihood ratio test is the lowest one for this operation, however not significant according to the P-value.[65]

Differences across and within industries

Nonlinearity in the data can be due to the fact that one does not sufficiently control for heterogeneity across and within an industry. Piece wise regressions is another way to test for the functional form of the relationship.

The use of piecewise regression analysis implicitly recognizes different functions fit to profits over varying ranges of marginal costs. Here, we differentiate across two lines: (i) between manufacturing and service; (ii) between small firms (=SME) and big firms (=BE).

Table 3.4 gives the results for PE. A number of interesting results can be distinguished. First,

[63] These parameters are transforming the variables in for instance: square, square root, log or reciprocal. Any transformed variable must be strictly positive.

[64] Other values are: θ = 2 means squared; θ = 1/2 means square root; and θ = – 1 means reciprocal.

[65] Lower values of the likelihood ratio mean that the observed result was much less likely to occur under H0 (null hypothesis) as compared to H1 (alternate).

Table 3.3 Box-Cox model: checking functional form of relationship for measuring PE

Model	θ	λ	θ or λ (LR statistic chi 2)			P-value		
	coeff	coeff	– 1	0	1	– 1	0	1
LHS only	– 0.034		31000000	3098.32	3000000	0.000	0.000	0.000
RHS only		– 0.5255	8578.44	3485.08	6633.4	0.000	0.000	0.000
λ		– 0.015	31000000	600.62	3000000	0.000	0.000	0.000
θ	0.343	– 0.025	31000000	4265.73	3000000	0.000	0.000	0.000

Note: The first row presents results of left-hand side (LHS) Box-Cox model; The second row gives results of right-hand side (RHS) Box-Cox model; The third row denotes the results for both sides Box-Cox model with the same parameters for LHS and RHS; The fourth row is identical to the third row but with different paramers for LHS and RHS.

Non-linearity revisited?

To test a simple form of non-linearity, we include a squared argument of AVC in our basic equation

$$\ln(\pi_{it}) = \alpha_i + \alpha_t - \beta_{1t}\ln(avc_{it}) + \beta_{2t}(\ln(avc_{it}))^2 + \varepsilon_{it} \tag{3.4}$$

Running equation 3.4, we compare the results with the preferred model at the aggregate level and use the following F-test to test whether the non-linear variable is significantly different from zero.

$$F = \frac{(R^2_{ext} - R^2_{base})/K)}{(1 - R^2_{base})(N - Z)} \tag{3.5}$$

Where K is the number of additional regressors, R^2 is the is square of the sample correlation coefficient (*i.e.* explained variance) with the underscores ext and $base$ referring to equation (3.4) and our basic model respectively, N is the number of observations and Z denotes the total number of regressors in the model.

The R^2_{ext} of equation (3.4) was slightly lower that the R^2_{base} of the basic model (*i.e.* 0.369 vs 0.376). Moreover, the F-test did not affirm that the non-linear variables are significantly different from zero. Hence, this result corresponds with the Ramsey Reset test, that the true specification of the preferred model is most likely loglinear.

differentiating between industries matters, pointing to the relevance of delineating the market (see for further analysis section 3.4.1). The PE-results for manufacturing are higher than for services. This is in line with earlier findings in chapter 2 and Creusen et al. (2006a,b). Second, apparently, differentiating between SMEs and BEs within manufacturing might be relevant too. In case of the services sector, difference between SME and BE are less pronounced. Hereafter, we control for differences across sector by including industry dummies when estimating PE at the aggregate level.

3.3.2 Fixed effects: impression of quantitative importance

In section 3.2.2, we argued that firm and time specific effects are important. Here, we examine whether these effects really matter in quantitative terms. We compare the results of the FE-model with a pooled OLS-model in two steps. In the first step, we estimate the following two equations at the industry level using pooled OLS and FE respectively

$$\ln(\pi_{it}) = -D_t\beta_t \ln(avc_{it}) + \mu_{it} \tag{3.6}$$

and

$$\ln(\pi_{it}) = -D_t\beta_t \ln(avc_{it}) + \alpha_i + \upsilon_{it} \tag{3.7}$$

The results of this first step provide evidence whether firm and time fixed effects are jointly important in quantitative terms. If β differs considerably in size between equation (3.6) and (3.7) then this may point to unobserved effects. This step does, however, not reveal whether either firm or time fixed effects are relevant separately in terms of quantity. Therefore, we need an additional, second, step and compare the results of our baseline model with the model without time dummies, $i.e.$ the difference between equation (3.3) and (3.7).

It turns out that the average (unweighted) PE-outcomes of the first step differ in size between equation (3.6) and (3.7) indicating that firm and/or time fixed effects are jointly important to some extent (see table 3.5). The average PE-level over time and across industries is in the case of the pooled loglinear OLS-model considerably higher than the PE based on an FE-model without time dummies. The number of negative outcomes of PE, which are economically unlikely, is larger for the OLS model. In contrast, the standard deviations of PE are somewhat larger for the FE-model. This higher standard deviations may be related to measurement issues that become more pronounced if one uses differences in time at the firm level. The issue of measurement errors is further addressed in section 3.4.2.

Unraveling the firm and time fixed effect, as the second step does, reveals that the time fixed effect seems to be less important in terms of size. The average value of PE hardly differs between PE with (see *PE FE + time*) and without (see *PE FE*) time dummies at the aggregate level (see table 3.5). Hence, the results for PE are less vulnerable for issues like

Table 3.4 Comparison FE-results at lower levels of aggregation, 1995-2006[a]

	Manufacturing[b]			Services[b]		
	Total	SME	BE	Total	SME	BE
	(1)	(2)	(3)	(4)	(5)	(6)
1995	-5.302***	-4.731***	-8.226***	-1.420***	-1.388***	-1.587***
	(0.18)	(0.21)	(0.28)	(0.020)	(0.021)	(0.060)
1996	-4.943***	-4.413***	-8.208***	-1.408***	-1.390***	-1.532***
	(0.15)	(0.15)	(0.30)	(0.020)	(0.021)	(0.063)
1997	-4.887***	-4.401***	-7.720***	-1.412***	-1.386***	-1.566***
	(0.13)	(0.14)	(0.39)	(0.021)	(0.023)	(0.064)
1998	-4.591***	-4.110***	-8.090***	-1.392***	-1.373***	-1.536***
	(0.20)	(0.20)	(0.27)	(0.020)	(0.020)	(0.068)
1999	-4.322***	-3.838***	-7.958***	-1.362***	-1.337***	-1.542***
	(0.15)	(0.15)	(0.30)	(0.020)	(0.021)	(0.069)
2000	-4.020***	-3.577***	-7.954***	-1.474***	-1.450***	-1.672***
	(0.11)	(0.10)	(0.39)	(0.019)	(0.020)	(0.063)
2001	-4.704***	-4.208***	-8.032***	-1.501***	-1.485***	-1.653***
	(0.12)	(0.12)	(0.42)	(0.019)	(0.019)	(0.063)
2002	-5.025***	-4.419***	-8.944***	-1.587***	-1.568***	-1.756***
	(0.15)	(0.14)	(0.63)	(0.020)	(0.021)	(0.063)
2003	-4.542***	-4.040***	-7.325***	-1.629***	-1.612***	-1.784***
	(0.53)	(0.56)	(0.99)	(0.020)	(0.021)	(0.063)
2004	-5.194***	-4.554***	-9.586***	-1.646***	-1.629***	-1.792***
	(0.39)	(0.39)	(0.33)	(0.020)	(0.021)	(0.064)
2005	-5.545***	-4.930***	-9.559***	-1.623***	-1.611***	-1.750***
	(0.19)	(0.19)	(0.41)	(0.020)	(0.021)	(0.063)
2006	-4.688***	-4.101***	-8.779***	-2.523***	-2.505***	-2.752***
	(0.29)	(0.28)	(0.41)	(0.13)	(0.13)	(0.57)
Time dummies	yes	yes	yes	yes	yes	yes
Industry dummies	yes	yes	yes	yes	yes	yes
Observations	87580	64766	22814	149948	127391	22557
Number of beid	22704	19166	3950	39893	36073	4268
R^2	0.35	0.35	0.46	0.33	0.33	0.32

[a] Robust standard errors in parentheses;*** p<0.01, ** p<0.05, * p<0.1; Industry dummies: 1 digit SIC-level.
[b] Manufacturing (maxsbi<5000); Services (maxsbi>5000).

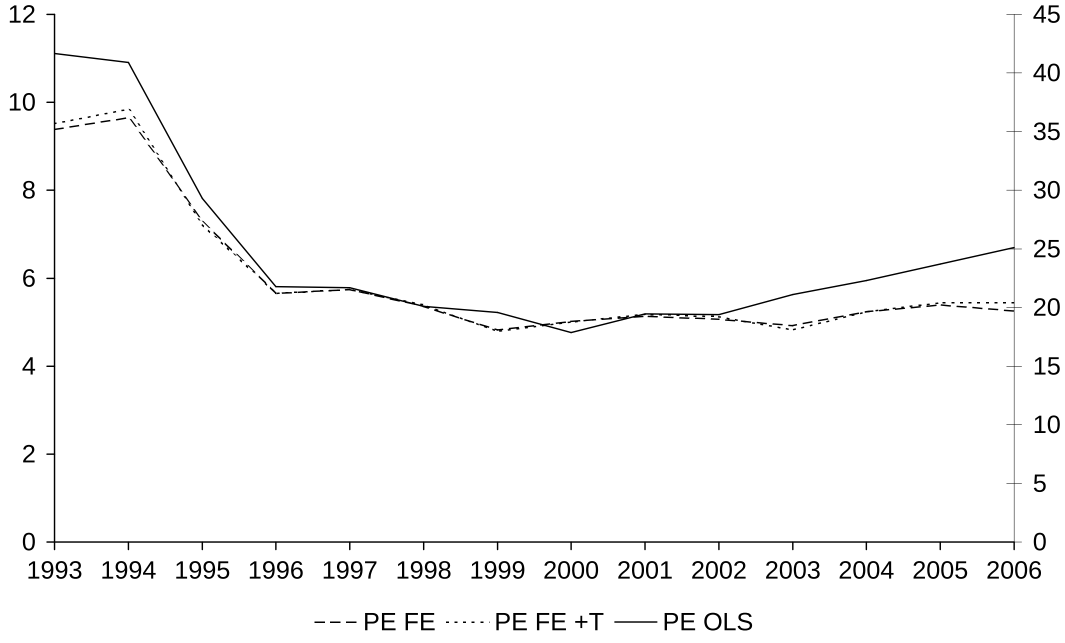
Table 3.5 Comparison results OLS and FE at 3 digit SIC-level, 1993-2006

Variable	Mean	Standard dev.	Minimum	Maximum	Observations
PE OLS	23.16	0.56	– 709	250	1949
PE FE	5.11	0.58	– 1108	60.2	1946
PE FE + time	4.62	0.79	– 1170	60.2	1949

Note that results are based on the cleaned data set.

business cycle or other exogenous shocks. And also the other statistics in table 3.5 indicate that there is close correlation between results of the baseline model and an FE-model without time dummies. This concerns not only levels but also changes per year, as can be seen from the development in figure 3.2.

Figure 3.2 Overall (unweighted) results OLS and FE, 1993-2006



Note: OLS-results on the right scale

The following summarizes this analysis. The potential effect on the level of PE of unobserved factors should not be ignored in quantitative terms. Our preliminary statistical assessment suggests that FE-model is "better" than the pooled OLS model (see also textbox below). Firm specific effects seem to dominate the difference between our preferred model and a pooled loglinear OLS-model. This finding may indicate that unobserved individual effects appear to be important and could be correlated with the explanatory variable. This will be further tested in the next two subsections.

Checks for relevance autocorrelation, industry and time dimension

Before running the four econometric techniques, we first check for possible autocorrelation with estimating the pooled OLS-regression at the aggregate level. Additionally, we include industry dummies and time dummies to control for differences across industries and across time and test whether those dummies are relevant.

Based on OLS, it turns out that the PE's (without any dummies) for the period 1995-2006 are strongly significant according to the high t-values. As we use time series, variables may exhibit autocorrelation, with the values in a given period depending on values of the same series in previous periods. Although the OLS-estimator is still unbiased if no other basic assumption is simultaneously violated, the variances of the parameter estimates will be affected. Consequently, the t-values will also be affected.[a] We, therefore, apply the Breusch-Godfrey test on autocorrelation.[b] This test is based on an auxiliary regression using the residuals from the original regression, regressed on a set of lagged residuals and all the variables which were used in the initial regression (see *e.g.*, Verbeek (2004)). In fact, we test whether the coefficients of the lagged residuals in the auxiliary regression are all zero, but that was not the case. The lagged residuals were strongly significant, hence suggesting that there is autocorrelation. Henceforth, we use the cluster option when needed, meaning that we take into account that within a cluster (in this case firms) the observations of AVC or profits are not independent over time.

We, then, run the FE model including time and industry dummies, and test there significance using the F-test. The R^2 in the extended model with time dummies slightly improves compared to the R^2 of the base model without time dummies. The F-test on the hypothesis that all additional time dummies have a zero coefficient is rejected. Moreover, including industry dummies further raise the R^2 and these dummies are also significant according to the F-test. Hereafter, we always use time dummies as well as industry dummies when needed.

[a] The t-values will be upwards biased in case of positive autocorrelation.

[b] This Breusch-Godfrey test is a more flexible test for the presence of first-order autocorrelation than the Durbin-Watson statistic. The former covers autocorrelation of higher orders and is applicable in situations where the regressors include lags of the dependent variable.

3.3.3 Fixed effects: correlation with alternative econometric techniques

Now, using more econometrics, we test whether the FE-approach produces different results with respect to PE than other econometric estimation techniques. We do that in the following way.

We compare the estimation results of PE at the industry level (*i.e.* 3-digit SIC-level) for all industries with three other econometric estimation procedures: (i) Pooled OLS, (ii) First Differences (=FD), (iii) Random Effect (=RE). This way we get an idea whether the FE-model is robust. In fact, we perform a number of regressions in which PE based on FE in market j at time t is explained through the results of PE of an alternative econometric estimator taking account of industry and time differences.

$$PE(FE)_{jt} = \gamma_j + \gamma_t + \gamma PE(X)_{jt} + \varepsilon_{jt} \tag{3.8}$$

where X is the PE of the other econometric technique, the $\gamma' s$ are parameters – with γ_t and γ_j being calendar year fixed effects and industry specific effects respectively[66] – and ε is an error term. A significant partial coefficient of $PE(X)$ suggests that there is a close correlation between the two different estimates in question and that the estimated PE is robust.

Table 3.6 shows that the estimation results for PE based on the FE-model significantly correlate with the results of the other econometric techniques using equation (3.8). Conditional on market characteristics (*i.e.* using time dummies and industry dummies), the size of the partial correlation coefficients is high and statistically significant across all estimation techniques.[67] The correlation coefficient is particularly high for FE and RE.

Table 3.6 Partial correlation coefficient between different econometric techniques for PE, 1995-2006

	FE	OLS	RE	FD
FE	1			
OLS	0.415	1		
RE	0.959	0.804	1	
FD	0.374	0.259	0.328	1

Note: The partial correlation coefficients are calculated with time and industry dummies.

[66] The calendar year fixed effects are included to take cyclical effects into account.

[67] We also eliminated all observations from the 1 and 99 percentiles reducing the impact of outliers. However, this does not affect the significance of the correlation.

3.3.4 Further econometric assessment

To gain more insights whether the FE estimator is to be preferred above another econometric estimator, we run additional econometrics tests. Using F-tests, t-tests and Hausman (1978), we examine and report whether there are significant differences between these estimators. We limit this analysis to the aggregate level and two other industries (*i.e.* 1510 and 7120) as we have already shown in the previous section that PE is robust using different econometric techniques at lower levels of aggregation. These additional econometric tests are only executed as an example for researchers interested. They prove whether one econometric technique is to be preferred above the other given the characteristics of our data set, and these can be unique and should always be tested by researchers.[68]

What is then the main difference between these econometric techniques considered? Whether one estimation method is more or less applicable than the other is basically related to the assumptions econometric techniques make for the firm-specific effect in relationship with the disturbance term (see equation 3.2). The four econometric techniques considered differ in this assumption. Questions that need to be addressed then are: if present, is the firm-specific effect correlated with the explanatory variable or with the disturbance term? Is the disturbance term autocorrelated?

Actually, a pooled OLS model assumes that the regressor (*i.e.* AVC) is uncorrelated with ε_{it} as in equation 3.2. In fact, OLS assumes that unobserved heterogeneity is not present, so u_{it} is IID. Put differently, if all the relevant regressors are included in the equation, then nothing is actually left as unobserved. In that case, the firm specific effect (α_i) drops out and OLS would suffice. The RE-model is based on the assumption that there is heterogeneity across firms. But, this techniques assume that unobservable individual effects are uncorrelated with the explanatory variable, and these individual effects are random. In contrast, the FE-model states that those effects may possibly be correlated with AVC. Finally, FD-estimation procedure is to a large extent similar to the FE-approach. Minor differences between the latter estimation techniques will be discussed below (see FE versus FD).

[68] Detailed results are available on request.

Relevance FE versus OLS

We test the relevance of FE-model by including the average of the firm specific effects to the OLS-specification (*i.e.* Mundlak specification test(Mundlak (1978))) and estimate this extended specification using the clustered standard error option. A significant coefficient means that the firm fixed effect is relevant. Indeed, we find that the coefficient of this additional variable is significant at the 1 percent level supporting the importance of firm heterogeneity. Moreover, we test the validation of our FE specification by using the (incremental) F-test. It improves. So, it turns out that the FE-model is better than OLS, and controlling for firm specific effects (α_i) is important.

Relevance FE versus FD

Also, FD assumes that the unobserved effect is correlated with the regressor. Similar to the FE estimation procedure, FD eliminates this unobserved effect. In contrast to FE, FD subtracts the observation for the previous year from the observation for the current year for all time periods.[69]

The choice between FE and FD as estimation techniques depends on the time series behavior of the error term (*i.e.* ε_{it}). If it is a white noise error term (*i.e.* serially uncorrelated) than the FE-model is more efficient and to be preferred. If the error term (without the firm specific effect) follows a random walk the FD estimation procedure is more efficient.[70] Essentially, FD and FE-models will be exactly the same when T=2, but if T >2 (as in our case) the two methods may differ. However, as long as strict exogeneity is present the FD estimation approach and FE-procedure should produce similar results for PE. If, however, the results of both methods are considerably different then this may generally point to problems of endogeneity or to omitted variable bias. Finally, FD is more vulnerable for outliers in the data than FE, because an outlier will affect two consecutive observations in FD while only one observation in FE.

[69] The FD method needs consecutive observations, it therefore loses at least one observation more for each firm compared with the FE-model.

[70] Basically, if ε_{it} is subject to AR (1) autocorrelation, taking first differences may approximately solve the problem of autocorrelation when α is close to 1 (in $\varepsilon_{it} = \alpha * \varepsilon_{it-1} + v_{it}$). Note that we control for autocorrelation in the FE approach as well.

We checked with a t-test whether the coefficients of both estimation procedures significantly differ.[71] It turned out that the coefficients of PE do not differ significantly for most years, and the developments over time were rather similar at the aggregate level and for the industries 1510 and 7120. As illustration, there was only a significant difference in competition level at the aggregate level in 1995 and 2006. This suggests that problems with endogeneity and omitted variables are limited, but we put these issues to a real test later on in subsection 3.3.5.

Relevance FE versus RE

As firm specific effects are present in our data set, we checked whether the individual effects are fixed or random. As already referred to, the main difference between FE and RE models lies in the role of unobserved individual effects (see *e.g.*, Hsiao (2003); Verbeek (2004)). Essentially, the RE model needs two conditions. First, the unobserved variable is randomly drawn from a given distribution. Second, this unobserved individual effect should be uncorrelated with the observable AVC in all time periods. In that case (α_i) may be treated as random variables.

The FE-model examines group differences in intercepts, assuming the same slopes and constant variance across firms within a market. However, when the observed variables are constant for each firm, a FE-model is not an effective tool because such variables cannot be included. In contrast, in a RE-model, the (firm) dummies act as an error term and the unobserved heterogeneity is not correlated with the explanatory variable. This type of model estimates variance components for firms and error term, assuming the same intercept and slopes. The difference among firms (or time periods) lies in the variance of the error term.
Apparently, RE is more attractive than FE because observed characteristics that remain constant for each firm are in the equation whereas they drop out in the FE model. But the RE-model is only consistent if there is no unobserved heterogeneity.[72]

[71] We employ the following t-test

$$t = \frac{coefficientFD - coefficientFE}{\sqrt{(\sigma_F D^2 + \sigma_F E^2)}}$$

[72] As OLS, RE puts the α_i in the error term, consequently observations may be correlated over time for this firm. This autocorrelation is, however, removed in the RE approach using Generalization Least Square Methods. In pooled OLS, this type of correction is not applied, hence the standard errors are biased.

To test whether the individual effects should be seen as fixed or random we apply a Hausman test (see Hausman (1978)). This Hausman specification test compares the FE versus RE-model under the null hypothesis that the individual effects (α_i) are uncorrelated with the other regressors (x_{it}) in the model.

$$H_0 : E(\alpha_i|X_i) = 0$$

or

$$H_1 : E(\alpha_i|X_i) \neq 0$$

If in our case α_i and c_i are correlated (*i.e.* H0 is rejected), an RE model produces biased estimators, violating one of the Gauss-Markov assumptions. So, an FE effect model is to be preferred. In fact, the PE based on RE-model is consistent and efficient under H_0, but inconsistent under H_1. In contrast, the PE based on FE-model is consistent but not efficient under H_0, and consistent under H_1.

Running the additional Hausman-test on the RE-model to test its relevance, we find in all three cases a P-value smaller than the significance level. Thus, the null hypothesis is rejected and FE estimation is preferred.[73]

Wrap up

To conclude, our preference for an FE-model is supported by different tests. As stated, whether one estimation method is more or less applicable than the other is basically related to the assumption for the firm specific effect in relation to the disturbance term. The tests above show that in our case the firm specific effect is correlated with the error term, supporting an FE-model.

3.3.5 Endogeneity problem

So far, we assumed that our explanatory variable (AVC) is exogenous and, for instance, not affected by the degree of competition. Suppose, however, this assumption of exogeneity is not true: what are the consequences for PE?

[73] Remember that the Mundlak test for testing the relevance of the OLS-model already signalled that there are significant firm specific effects.

The problem of endogeneity arises when an explanatory variable is correlated with the error term in our model. In fact, the endogeneity bias in PE is the correlation between the endogenous variable, in our case AVC, and the error term as the regressors should be uncorrelated with the error terms (*i.e.* $cov(AVC, \varepsilon) = 0$) to reflect the true parameter without a bias. Endogeneity may occur when there are omitted variables, measurement errors in right-hand-side variable or simultaneity (see *e.g.*, Verbeek (2004)).

Hence, if AVC turns out to be endogenous in practice, the FE-estimation of PE is likely to be biased and inconsistent if the endogeneity problem is not properly addressed. More precisely, the variation in profits does not arise solely from variation in AVC in that case.

Despite the fact that earlier we concluded that endogeneity (and omitted variables) problems might be limited according to a comparison of the results of FE and FD, it still can be argued that AVC is not exogenous in practice for two reasons. First, the way we define profits and AVC can generate endogeneity problems.[74] Second, AVC could also be affected by earlier profits (reverse causality) or unobserved variables may affect profits as well. For instance, innovation might affect the marginal costs, whilst innovation depends on competition. Also, if a firm on product market X has relatively more buyer power on its input markets, for instance, on the wholesale market or labor market than its competitors have, marginal costs are no longer exogenous.[75] Hence, the situation on these input markets will have consequences for the marginal costs. Market power on product markets and buyer power on input markets are then interrelated (see, inter alia, Creusen et al. (2008)).

To cope with endogeneity, we need estimates of the endogenous variables expressed exclusively in terms of exogenous factors. In general, the literature (see *e.g.*, Cameron and Trivedi (2005); Wooldridge (2002)) provides two approaches to deal with this issue: use time lags (see textbox) or instrumental variables.

[74] Because profits defined as $p_i x_i - c_i x_i$ are the dependent variable while $\frac{c_i x_i}{p_i x_i}$ is the explanatory variable, one may worry that the estimate of β is biased. Moreover, if, say, revenue $p_i x_i$ is observed with an error ρ_i, then a negative relation between $\rho_i p_i x_i - c_i x_i$ and $\frac{c_i x_i}{\rho_i p_i x_i}$ is induced even if profits and costs would be unrelated in the market.

[75] Boone et al. (2010a) shows that although competition may affect (marginal) costs, one can still interpret changes in PE as changes in intensity of competition.

Endogeneity and using lags

To tackle the problem of endogeneity, one standard option is lagging the potentially endogenous explanatory variable. This option relies on the assumption that the lagged value of the explanatory variable is uncorrelated with the error term of equation (3.3). The argument is that although the current values of the AVC might be endogenous to profits, it is unlikely that past values of AVC are subject to the same problem.

Running this strategy at the aggregate level, the coefficients for the one-year lagged AVC are all significant and positive (absolute value). This underlines that efficiency differences between firms are relevant in explaining differences in profits. Moreover, these results significantly correlate with the other estimation procedures, but correlation coefficients are low. Note that this strategy of using lags provides no way of gauging empirically how serious the endogeneity problem is, and whether the solution adequately deals with it. It only gives an alternative approach and a better option is available as discussed below.

Instrumental Variables (=IV)

An informative way to test and control for the endogeneity problem is using instruments. Such instruments require the following two important criteria. First, each instrument should be exogenous in our baseline model. This means that a good instrument is uncorrelated with the error term (ε_{it}), which gives consistent PE. Second, the instrument should be of relevance. This implies that the instrument should be highly correlated with AVC, which gives accuracy to our estimator (Heckman (1997, 2008)).

Broadly speaking, coping with endogeneity, Cameron and Trivedi (2005) and Wooldridge (2002), for instance, show that one should estimate two regressions. First, one runs an auxiliary regression (*i.e.* equation (3.9)) that determines the explanatory variable, in our case AVC, using instruments. Then, include this residual (or the predicted value) of equation (3.9) in the main equation (3.10). Significance of the error term (or the predicted value) in the main regression points to endogeneity of that variable.

$$\ln(avc_{it}) = \gamma_i + \gamma_t - \delta \ln(iv_{it}) + \eta_{it} \tag{3.9}$$

$$\ln(\pi_{it}) = \alpha_i + D_t - \beta_t \ln(avc_{it}) + \zeta_{it}\hat{\eta}_{it} + \varepsilon_{it} \tag{3.10}$$

We know from the empirical literature that good valid instruments are hard to find in practice. In that respect, our data set is a good example. For instance, often factor prices like oil prices are used as instruments, but those prices are not in PS and hence not in our data set.

Here, we use the depreciation ratio (= depreciation expenditures over revenues) and size class as instruments. To get an idea how relevant those instruments are, we run a simple

Two tests for relevance IV instruments

Researchers that use IV have two types of econometrics tests at their disposal for testing the relevance of instruments for coping with endogeneity: (i) tests for validity of instruments, (ii) tests for exogenous of explanatory variables.

Both the *Sargan test* and *Basmann test* test the instruments validity and model specification jointly. These tests are only possible if one has more instruments than potentially endogenous variables. Sargan and Basmann are tests of over-identifying restrictions. A rejection of their null hypothesis casts doubt on the validity of the instruments.

Whether or not the right-hand-side variables are weakly exogeneous can be tested by the *Wu-Hausman test* and *Durbin-Wu-Hausman tests*. The null hypothesis indicates that the OLS estimator of the basic equation would yield consistent estimates. If the null hypothesis is not rejected then FE is consistent and efficient. IV is also consistent but inefficient. In contrast, a rejection of the null hypothesis indicates that endogenous regressors' effects on the estimates are meaningful, and IV is required. More precisely, a small p-value indicates that the null hypothesis is rejected and that OLS (or FE) is not consistent.

first stage regression including the instruments leaving out AVC as potential endogenous regressor. We checked the relevance of those instruments by looking at the F-value whether they are significantly different from zero. The F-value was much larger than 10 pointing to the fact that the instruments are valid (rule of thumb). Then, taking account of those instruments, we run an IV-regression at the aggregate level and for our two randomly chosen industries (*i.e.* 1510 and 7120) to get an idea about the relevance of the instruments.

The results for IV at the aggregate level – where we estimate equation (3.9) and equation (3.10) across all countries – come close to the one from the FE-model (see table 3.7). PE is often slightly higher using IV, except for 2006. The difference between IV and FE are more pronounced for industry 1510 and 7120. What does this mean? The IV-results seems less plausible with huge changes over time. Looking at the results for the Sargan and Basmann tests to check the validity of our instruments, the p-values of these two tests are for all three examples lower than 0.05. This means that the chosen instruments are not valid. Stated otherwise, the null hypothesis is rejected. Although, the null hypotheses of the two Hausman tests for the endogeneity bias are rejected, the results of these tests are not reliable as our chosen instruments are not valid.

Despite the fact that endogeneity might be an issue, we conclude that we do not have valid instruments to cope with this potential bias appropriately.

Table 3.7 Endogeneity results: Overall and two industries, 1995-2006

	(1)	(2)	(3)	(4)	(5)	(6)
	Overall		1510		7120	
	FE	IV	FE	IV	FE	IV
1995	-1.726***	-1.698***	-5.291***	14.10	-1.508*	– 0.483
	(0.020)	(0.054)	(0.58)	(11.0)	(0.76)	(1.16)
1996	-1.724***	-2.635***	-5.086***	11.42**	-2.246***	– 1.323
	(0.020)	(0.064)	(0.65)	(4.42)	(0.76)	(1.17)
1997	-1.717***	-2.381***	-6.083***	8.023	-2.225***	– 1.642
	(0.020)	(0.061)	(0.70)	(5.04)	(0.78)	(1.21)
1998	-1.727***	-2.165***	-5.651***	7.662*	-1.962***	– 2.226
	(0.020)	(0.069)	(0.64)	(4.42)	(0.67)	(1.60)
1999	-1.686***	-2.338***	-5.604***	35.86**	-2.186***	-1.431*
	(0.021)	(0.071)	(0.53)	(16.8)	(0.62)	(0.78)
2000	-1.783***	-7.474***	-3.198***	-51.13***	-1.011***	-21.20***
	(0.020)	(0.096)	(0.52)	(9.33)	(0.12)	(3.20)
2001	-1.823***	-2.238***	-5.581***	18.40**	-2.217***	-5.705**
	(0.020)	(0.054)	(1.03)	(8.93)	(0.77)	(2.57)
2002	-1.906***	-2.229***	-8.421***	29.27***	-2.157***	– 3.049
	(0.021)	(0.057)	(0.90)	(9.71)	(0.38)	(3.19)
2003	-1.968***	-2.252***	-10.12***	11.91**	-1.656***	-4.393**
	(0.022)	(0.062)	(0.91)	(6.03)	(0.32)	(1.74)
2004	-1.981***	-2.247***	-6.705***	27.64*	-2.078***	-3.916***
	(0.022)	(0.068)	(0.71)	(16.2)	(0.36)	(1.16)
2005	-1.963***	-2.091***	-4.555***	56.73	-1.585***	-2.985***
	(0.022)	(0.068)	(1.10)	(125)	(0.30)	(0.75)
2006	-3.015***	15.44***	-10.20***	20.20*	0	0
	(0.11)	(1.08)	(1.30)	(10.3)	(0)	(0)
Time dummies	yes	yes	yes	yes	yes	yes
Industry dummies	yes	yes	no	no	no	no
Observations	250194	231268	1728	1660	272	265
R-squared	0.27		0.35		0.56	
Sargan test	0.000	0.000	0.000	0.000	0.000	0.000
Basmann test	0.000	0.000	0.000	0.000	0.000	0.000
Wu-Hausman	0.000	0.000	0.000	0.000	0.000	0.000
Durbin-Wu-Hausman test	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

3.4 Robustness checks

Having argued why we prefer to estimate PE with an FE-model, we now turn to a number of alternative estimations to examine the robustness of the estimation results of our basic equation (3.3) in terms of statistical issues. More precisely, this section addresses the potential impact of the following three issues on the outcome of PE related to the quality of the panel data: (i) defining the relevant market (see section 3.4.1), (ii) measurement issues (see section 3.4.2 and section 3.4.3) and (iii) selectivity problems (see section 3.4.4). As in section 3.3.3, we use equation (3.8) here to test the robustness of PE in these three situations.

3.4.1 Defining the relevant market

As argued in section 3.2.4, it is difficult in practice to delineate the relevant market. Hence, the results of PE can be vulnerable to this issue.[76]

To provide an impression of the impact on PE of defining the relevant markets in terms of the digit level SIC-code, we calculate the PE at the two and three digit SIC-code respectively. Then we aggregate the results from the three digit level to the two digit level, and compare these (aggregate) results with the findings directly based on estimations at the two digit level.

Applying equation (3.8) there exists a strongly significant correlation between these two measures of competition. The partial correlation coefficient is 0.646 and statistically significant with a t-value of 24.83. This outcome suggests that in case of aggregation the results contain relevant information on the development of competition in underlying submarkets. On average, the results based on aggregates correspond with the more detailed information. Hence, even if the (relevant) market is not correctly defined, PE is still informative about competition on the submarkets included in the defined market.

3.4.2 Measurement issues of PE

Firm level data based on surveys are prone to measurement errors for a number of reasons. Hence, also our two key variables (*i.e.* variable profits and AVC) could be mismeasured.

[76] Note that for traditional measures this is also an issue (see Creusen et al. (2006a). Moreover, concentration rates like Herfindahl need figures for other countries as well to measure competition correctly.

For instance, measurement errors in either our profit measure or in our cost measure could be due to faulty response in the surveys as a consequence of badly designed questionnaires, respondent bias including deliberate distortion of responses (*e.g.* strategic behavior), and processing errors.

Here, we test how measurement errors affect the variables we are interested in, including our competition measure PE. Measurement errors can be grouped into three types: unobserved effects, systematic errors, and random errors.

First, as already discussed, unobserved effects are taken into account in FE. As long as these unobserved effects are constant over time and firms, they pose no problem. Similarly, systematic measurement errors in both profits and AVC are not an issue if they are constant over time. Those errors cancel out in the FE-model as each observation per firm is compared with the mean of this observation over time that also includes the unchanged measurement error.[77] As shown above, systematic over- or underestimation of profits and costs per firm is picked up by the firm fixed effects in the regressions. For instance, a firm has relatively higher costs than its competitor due to producing a better product in terms of quality, but this quality is not observed by us in terms of higher nominal values of output as we have in our dataset. Then this bias is properly dealt with in FE as long as a firm's quality (being the measurement error) is constant over time. However, random errors (in AVC) in our basic equation may give problems as shown below.

Random errors in profits

A random measurement error in the dependent variable profits is not severe for estimating PE with an FE-model as this error is taken into account in this econometric model. To illustrate, assume we have the following[78]

$$y = \pi + \omega \tag{3.11}$$

where π is the true value for profits, ω represents the measurement error in profits, and y is

[77] For instance, we denote the observed profit level for firm i at time t by $\pi_{it} u_i$. Hence the observation error equals $(u_i - 1)\pi_{it}$. Similarly, the observed marginal costs are denoted by $c_{it} v_i$. The assumption that we make is that these observation errors may differ between firms but are constant over time for the years. These errors disappear when taking differences.

[78] For ease of notation, we neglect i, t and 'logs'.

the measured or observed profits. Suppose our real model is

$$\pi = \alpha - \beta mc + \varepsilon \tag{3.12}$$

where mc denotes the true value for efficiency. We assume that the error, ω, has zero mean, $E[\omega] = 0$, and $V(\omega) = {\sigma_\omega}^2$. Moreover, this error is uncorrelated with the costs, $cov(\omega, mc) = 0$, and also uncorrelated with the true value for profits, $cov(\omega, \pi) = 0$. Rewriting equation (3.12) in terms of observed variables by substituting for π from equation (3.11), we get

$$y = \alpha - \beta mc + \varepsilon + \omega \tag{3.13}$$

Hence, there will be no bias in PE if only profits are measured with random error and our cost variable is measured without random error, since ε and ω are both random and not correlated with AVC. However, it gives larger standard errors in PE due to larger variation.

One comment put forward in measuring PE is that this competition measure seemingly relies conceptually on economic profits, whereas the available data in the PS is based on accounting profits. For instance, Fisher (1987) and Fisher and McGowan (1983) argue that different (fiscal) accounting rules between firms can result in differing accounting profits that bear no obvious structural relationship to economic profits. But as long as our profit measure is a proxy with random errors independently of the underlying true value of profits, this comment does not matter as can be seen from equation (3.11) to (3.13). This is also the case when the difference in profits between both concepts are constant over time.

Random errors in AVC

The consequences for PE are different if there exists a random measurement error in AVC: the explanatory variable. Then a (downward) bias in PE exists. Assume we have the following

$$AVC = mc + \mu \tag{3.14}$$

where μ denotes the random error in AVC, with the same conditions as discussed above for ω. Again, rewriting equation (3.12) in terms of observed costs by substituting for mc from equation (3.14)

$$\pi = \alpha - \beta AVC + \kappa \tag{3.15}$$

where

$$\kappa = \varepsilon + \beta\mu$$

Thus β (or PE) is also related to κ, since $cov(\kappa, AVC) \neq 0$. Therefore, OLS provides biased estimates in levels in case of random error in AVC, since κ is not distributed independently of AVC. [79] β is underestimated as one can show that

$$plim\hat{\beta} = \frac{\beta}{(1+\varphi)} \tag{3.16}$$

where $\varphi = \sigma_{\mu}^{2}/\sigma_{mc}^{2}$. As the denominator is larger than one, OLS will underestimate PE (see Maddala (2001); Verbeek (2004)). The degree of underestimation depends on φ, also referred to as the noise-to-signal ratio, representing the variance of the measurement error (the noise) in relation to the variance of the true value (the signal). If this ratio is high then the bias in PE is large.

The analysis implies that we should be careful in interpreting the level of the estimated β's. The bigger the variance in random errors μ in relation to the variance in mc the more the applied econometric technique underestimates β. However, as long as the variance of μ is constant over time in a given sector, we can interpret changes in $\hat{\beta}$ over time as changes in the underlying competition intensity (*i.e.* β).[80]

Finally, it is sometimes argued that FE-models control for these kind of errors better than other econometric techniques like OLS. This statement is wrong. Although FE eliminates one source of bias by differencing out individual effects, it creates another bias (*i.e.* attenuation bias due to smaller sample with larger variability of avc) that may result in even more inconsistent estimates than OLS (see Hsiao (2003)).[81] But compared to the FD-approach,

[79] Here, for ease of notation we refer to OLS, but other econometric techniques lead to measurement errors in the same way as well.

[80] One could analyze this noise-to-signal ratio further by making some assumptions about the error variances. We will not undertake this type of analysis here as it is beyond the scope of this chapter. We are more interested in the correlation between the direct regression and alternative regressions coping with measurement issues. Moreover, the results of PCM might be bias in the same vein.

[81] In fact, Hsiao (2003) states that if the serial correlation of μ is less than that of mc ("...as seems often likely to be the case"), using FE increases the noise-to-signal ratio.

the measurement error bias might be smaller with an FE-model (see Griliches and Hausman (1986)).

3.4.3 Testing for measurement issues

We apply a number of alternative options to gauge the relevance of measurement issues on PE:[82]

- Reverse regression technique
- Other definitions for profits and costs
- Benchmarking

Below, we first explain what these options aim to do. Then the results of these options are discussed and presented in table 3.8.

Reverse regression

The first alternative is that we change the dependent variable and the explanatory variable, in the literature referred to as reverse regression (see *e.g.*, Maddala (2001); Schaefer and Visser (2003)).[83] Although we assume that AVC (*i.e.* "cause") affects the profits (*i.e.* "effect") and not the other way round, this reverse regression approach is one way to test the impact of measurement issues on PE. Particularly, whether both variables are measured with error. We estimate the following relationship using FE.

$$\ln(avc_{it}) = -\gamma_t \ln(\pi_{it}) + \alpha_i + D_t + \varepsilon_{it} \tag{3.17}$$

Normally, if measurement problems (or missing variables) are not a serious issue we would expect that

$$1/\beta_t = \gamma_t \tag{3.18}$$

[82] Note that the measurement problem could also be taken care of by using IV estimation procedures (see *e.g.*, Maddala (2001) and Verbeek (2004)). We already discussed IV in section 3.3.5.

[83] One could also use orthogonal regression techniques here (see *e.g.*, Schaefer and Visser (2003)). This technique can be applied if both dependent and independent variables include measurement errors or when the direction of causation is unknown. This is beyond the scope of this chapter and we leave it for further research.

Other definitions for profits and AVC

Ideally, our indicator for the regressor should rank firms according to their efficiency. One could argue that the marginal cost concept fulfills this requirement in the best way, since this concept includes all costs varying with small changes in production: firms that are more efficient will have lower costs than inefficient firms when they change the volume of their production. However, the available data set does not allow us to measure or to estimate marginal costs. Therefore, we use AVC as a proxy for marginal costs.

The question arises whether the AVC is a good proxy for our marginal cost concept (see also textbox *Differences between AVC and marginal costs*). Stated otherwise: are the observed costs included in the AVC really related to efficiency? We depart from our AVC concept to gauge the relevance of the cost concept in three ways: (i) adjust the cost concept for scale economies, (ii) use another efficiency measure, (iii) use other cost (and output) definition.[84]

Differences between AVC and marginal costs

Assume that the marginal costs is the true indicator for measuring efficiency but unfortunately not measurable given our data set. Additionally, the values of AVC that we observe as proxy differ randomly from the true values of the marginal costs. From economic textbooks we know that the AVC is equal to marginal costs if there are no economies of scales. But if economies of scale are present then marginal costs is not equal to AVC.

$$AVC_i = MC_i - X\frac{\partial C}{\partial X}$$

Hence, if those scale economies are random, our estimated PE is downward biased compared to the results based on the non-observed marginal costs as can be derived from equation (3.16).

Economies of scale lead to differences in costs between firms. There is a rather simple way to control our cost measure for economies of scale following the idea of Bikker and Van Leuvesteijn (2005). First, they approximate marginal costs as the dependent variable by estimating a regression using production and production squared as explanatory variables.[85] The

[84] An alternative, but rather comprehensive, way to obtain another efficiency measure for each firm is to apply a frontier type of analysis based on Data Envelope Analysis. The efficiency score of each firm can be directly used in our basic equation. We leave this alternative approach for further research.

[85] Bikker and Van Leuvesteijn (2005) also apply a translog cost function to control for scale economies. That option is not possible for us due to a lack of factor prices.

residuals of this equation can be interpreted as (marginal) costs adjusted for scale economies. Then, these corrected costs are used as explanatory variable in equation (3.3).

Second, an alternative indicator in our data set that can serve as a proxy for the efficiency level of a firm is its (labor) productivity level. The data set allows us to calculate productivity levels (*i.e.* gross output per employee) at the firm level.[86] Clearly, the sign of the coefficient of this indicator should be the opposite from the one for AVC. A higher productivity level means higher efficiency.

Finally, we use both for the (variable) profits and for the AVC broader concepts to get an idea about the relevance of our lean concept for both variables in terms of measurement issues. In these broader concepts (see textbox *Different definitions for profits and costs possible* in section 3.2.3), we also include the costs and profits of other activities. We consider those issues as we deliberately want to introduce measurement errors, here defined as the difference between the lean and wide definition (*i.e.* net turnover of other activities and values of purchased commodities). As in a multiproduct setting, the allocation of (common) costs to the various products produced influences the profit margins on the products in the market(s) for which the PE-measure is estimated.

Benchmark firm

Picking up the original idea of Boone (2000a) that one should use some kind of benchmark firm when measuring PE is one way to get an idea of the importance of random errors in AVC. The intention of this benchmark option is that it tries to control for definition and conceptual problems by standardizing the profit and AVC of each firm to some kind of benchmark in its market. This might mitigate the potential measurement errors if they occur across firms in an industry.

We denote the profits and costs at time t of this benchmark firm by resp. $\bar{\pi}_t, \bar{c}_t$. The benchmark firm could be the median firm or the least efficient firm in the market. The exact identity of this firm does not matter as it is only used as an additional control for measurement issues next to the firm-fixed effect. Here, we relate each observation of a firm to the median

[86] Note that the labor productivity is in nominal levels and not adjusted for inflation over time.

outcome at the industry level for every year.

Hence, we estimate the following relationship for this variant[87]

$$\ln\left(\frac{\pi_{it}}{\bar{\pi}_t}\right) = -D_t\beta_t ln\left(\frac{avc_{it}}{a\bar{v}c_t}\right) + \alpha_i + D_t + \varepsilon_{it} \tag{3.19}$$

Results measurement issues: correlation between different variants

Using equation (3.8), table 3.8 presents the correlation coefficients of our analysis of measurement issues at the 3-digit level for all variants considered. It shows that our concepts for profits and efficiency using FE-model significantly correlates with all the variants we considered: (i) reverse regression (=RR), (ii) adjusted marginal cost concept (=Mc adj),(iii) labor productivity (=LP), (iv) other definitions (=Other) and (v) benchmark firm (=Median). Apparently, measurement issues are not severe and appear not to bias the PE in that respect.

As one would expect the results of RR negatively correlate with FE due to $\gamma = 1/\beta$. The latter is also reflected in the size of the coefficient being larger than –1. The partial correlation coefficients of the other variants are positive and statistically significant. Particularly, the results for PE based on LP correspond closely to the one based on FE, indicating that labor productivity is a very likely candidate to replace AVC to capture efficiency in circumstances were researchers cannot calculate AVC but labor productivity. Moreover, it indicates that the interdependency between profits and costs due to their definition is most likely not a problem for measuring PE.

3.4.4 Selectivity issues

The PE-outcomes are based on a sample of firms of the Dutch economy from subsequent waves of the PS. This data set is an unbalanced panel where we do not have all yearly observations for every firm in the population over the period 1993-2006 (see figure 3.1). In general, unbalanced panel data are not a problem for statistical packages. However, incomplete panel data can be a problem for competition analysis in case of sample selection bias. Sample selection leads to inconsistency of the PE parameter if the missing observations are

[87] If we rewrite this equation, it differs from equation (3.3) as it also includes the following terms:

$$\ln\bar{\pi}_t + D_t\beta_t \ln a\bar{v}c_t$$

Table 3.8 Measurement issues: partial correlation coefficient between different variants for PE, 1995-2006

	FE	RR	Mc adj	LP	Other	Median
FE	1					
RR	– 4.356	1				
Mc adj	0.719	**– 0.000**	1			
LP	0.934	0.018	0.663	1		
Other	0.440	**0.001**	0.268	0.066	1	
Median	0.607	– 0.004	0.381	0.067	0.707	1

Note: if bold then the result is not significant at the 5 percent confidence level (of p-value).

related to the relationship of profits and efficiency we want to examine. In other words, the observations we employ may not be randomly drawn from the population (see *e.g.*, Hausman and Wise (1977); Heckman (1979)). Those firms may significantly differ in this relationship.

Here, we examine what the impact of selection bias might be on the robustness of the estimation results of PE. Note that selectivity in the independent variable (exogenous), in our case the AVC, is not directly a problem, whereas selectivity in the dependent (endogenous) variable, i.e. the variable profits, can be a problem. This is due to the fact that the latter may also include a non-random error component and this may directly affect the slope of β, and hence PE, while a (non random) error component in costs will probably only affect the t-statistics of PE.

In our data set, selectivity problems might arise due to, for instance, truncation of the sample or non random non-response. Truncation occurs because small firms in the manufacturing industry are missing as these firms are not present in the PS. Or, if the non-response of firms is directly related to their profit performance.

Does this matter and how to deal with selectivity issues? To a large extent, as it is out of our control, we cannot reconsider the panel data set as a consequence of the sampling methods of Statistics Netherlands. Also, selectivity issues related to non-response and confidentiality requirements we cannot control let alone correct for because these issues are under supervision of Statistics Netherlands. However, we can go the other way round if we take a sample out of what we have at our disposal to test for selectivity issues. Moreover, there are cases where we ourselves deliberately removed observations out of our panel data set that may bias the results for PE. We can include them again. Hereafter, we present three variants taking account of these selectivity issues.[88] We first explicate what the aim of these variants is. Then the results are presented in table 3.9 and discussed below.

Variant I: Issue of balanced vs non balanced results

Although panel data sets allow researchers to apply different econometric techniques, subsequent waves of panels are subject to nonresponse and consequences of sampling method. These items are related to for instance resistance of firms to fill in the questionnaire each year and small firms not being sampled each year due to the way Statistical Offices conduct surveys. Consequently, firms drop out of the sample in subsequent waves. The subsample with complete data for all waves of the panel becomes smaller and smaller over time, and the remaining sample gradually becomes less representative for the population than the original sample. This is called the attrition problem of panel data and this type of missing data may generate biassed parameters (see *e.g.*, Hausman and Wise (1979); Ridder (1990); Verbeek and Nijman (1992)).

The issue at stake is that there might be self selection problems. If the probability of attrition is correlated with the expected response, then estimation techniques produce biased and inconsistent estimators. Fortunately, Statistics Netherlands carries out a so called refreshment sample that helps to mitigate the effects of attrition (see *e.g.*, Hirano et al. (2001)). It adds new firms randomly sampled from the original population. Nonetheless, this only partly solves the problem, because small firms may not be interviewed in two consecutive years.

As discussed, in theory it is not necessary to observe all the firms in the market for measuring PE. Since PE measures the degree of competition by comparing at least three firms with different efficiency levels, an unbalanced panel data set will do to measure PE in practice. However, the availability of few observations due to small samples may bias PE as the probability that the estimated coefficient is more likely to be different from the actual (or true) value of PE than in case of large samples (see *e.g.*, Verbeek (2004) and the BLUE-conditions).

[88] In our basic model, we applied an unweighted fixed effect regression technique. Hence, firms are not weighted according to their importance for a market. A weighting technique is often applied in case of serious sample problems. In an economic sense, we already checked for heteroskedasticity problems by taking into account that the variance of the residual may differ between groups of firms.

To figure out the relevance of the attrition problem, we estimate the baseline specification using a panel based on more consecutive observations of firms than our main data set. We limit this approach to two selection criteria to have sufficient observations for more firms to generate enough variation in our key variables: firms with at least 5 observations, respectively 9 observations. Huge differences between the results of those panels and the PE of our baseline model may suggest that attrition is an issue.

Variant II: Other subjective selection issues

Besides ignoring loss-making firms, we performed several other 'cleansing' activities in section 3.2. These activities could lead to biased parameters for PE.

In this option, we construct an adjusted cleaned data set including those observations from the three cleansing steps that were initially removed from the cleaned data set that have been used for the FE-results so far.[89] In this alternative data set, we include: (i) same observations for firms in two consecutive years; (ii) huge changes in output or employment; (iii) negative intermediate inputs.

Variant III: Issue of neglecting loss-making firms

The advantages of the logarithmic specification of the baseline model is that the β of equation (3.3) can be directly interpreted as an elasticity. Moreover, transformation of variables in logs normally reduces heteroskedasticity mitigating the variance of variables between firms.[90] However, due to the log transformation in our baseline model, all firms with losses are ignored reducing the number of observations. But is this a problem? Gives this subgroup of the sample an unprecise picture of the entire sample, let alone the entire population?

According to Boone (2008) it does not matter for PE when firms with losses are ignored because their relevance for the competitive pressure is still reflected in the profits of the remaining or observed firms. Stated otherwise, from a theoretical perspective the slope of PE is not biased due to censoring of only the positive observations for profits. But Boone (2008) considers a two-stage game in a static setting[91], then the issue of incumbents making losses

[89] Note that the other cleansing steps are still executed including removing loss making firms.

[90] Using logs rather than levels reduces the dispersion of values that vary with different levels of the fitted value of the dependent variable.

[91] In the first stage, a firm enters the market if its expected profits is larger than the entry costs. In the second stage,

does not arise as those firms would choose not to produce at all. However, in a more dynamic setting, it can be argued that it is important to take into account the existence of firms with losses as well. In that case, firms weight the cost of operating at a loss in the current period against the alternative of leaving the market now and with the option to re-enter it later on (see Maliranta et al. (2007)). Actually, there are firms that may (temporarily) have negative performance in terms of profits, but nonetheless succeed in surviving. Looking at our data set, the number of observations we do not observe is limited. As we can infer from the cleansing operation in section 3.2, approximately 15 percent (= 59143 / 387575) of the observations in our panel data set were firms that were making no profits in the period 1993-2006.[92] Their variable profits were negative and those firms were removed from the cleaned data set.

Although we talk about a small fraction of our data set, we assess one variant to gauge how relevant the issue of neglecting those loss-making firms is. This variant rescales the negative profits of firms, in such a way that the negative profits in our sample disappear in the observed period.[93] We transform the profits of all firms with losses to zero (in logs). Then, we include an additional dummy to our basic equation (3.3) with the value one in a particular year for firms with losses and zero for the remaining firms.

Table 3.9 Selectivity: partial correlation coefficient between different variants for PE, 1995-2006

	FE	BAL5	BAL9	SUB	SEL
FE	1				
BAL5	0.475	1			
BAL9	0.226	0.357	1		
SUB	0.976	0.858	0.917	1	
SEL	0.332	0.347	0.399	0.328	1

all firms that entered choose simultaneously and independently their input and output factors.

92 Note that we refer to observations, hence a firm that has negative profits in, for instance, two year counts for two observations.

93 Another option would be to estimate our preferred model without using logs. In that case, our sample can be extended including those firms operating at a loss. However, this option is not informative anymore, as we already concluded in section 3.3 that model using "logs" is preferable as functional form. Anyway, a recent Finnish study on measuring competition find that the two regressions are strongly correlated in levels, and changes over time cohere largely (see Maliranta et al. (2007)). This implies that ignoring loss-making firms is seemingly not generating a bias in the estimates of PE.

Results selectivity issues: correlation between different variants

Table 3.9 summarizes the three variants coping with selectivity issues by looking at the correlation coefficients from equation (3.8) at the 3-digit SIC-level. Particularly, the partial correlation coefficient is very high and significant for FE with variant II with subjectivity (see column SUB, in table 3.9). It clearly shows that those 'other cleansing' activities hardly matter since the coefficients for PE of variant II are almost exactly the same as the one for the FE-model.

The remaining two variants – balanced panel with at least 5 or 9 observations (*i.e.* $bal5$ and $bal9$ respectively) and rescale (=SEL) – also correlate significantly with FE. The correlation coefficient between FE and $bal5$ is somewhat larger than between FE and $bal9$, but still $bal9$ is significantly correlated with FE. Hence, the attrition problem appears to be limited. Statistics Netherlands recently concluded that there can be huge differences between a balanced and unbalanced panel when they compared aggregated firm level results with National Accounts for the chemical industry (Van Leeuwen et al. (2008)). However, for measuring PE, probably this is not a problem. According to the significant correlation results between FE and SEL, the consequences of ignoring loss-making firms seem to be limited as well.

3.4.5 Further research issues

Although the consequences of selectivity issues appear to be limited as presented in table 3.9, there is a more sophisticated option available to test the importance of excluding loss making firms for estimating PE. The idea is to extend the baseline model with an equation that 'determines' making losses (see Hausman and Wise (1979)). If the error term of this (second) equation correlates with the error term of the baseline model than there exists selective non-response.

This option is often applied in labor economics (see *e.g.*, Heckman (1979) and Heckman et al. (1999)), where one first determines whether or not somebody participates in the labor market and, thereafter, the wage level in case someone participates. In a similar way, this option can be used in case of innovation. First, one models whether or not a firm innovates, and then, one models the amount of innovation expenditures (see Creusen et al. (2006c) for an example).

In our case, however, the argument for applying such a selection model is not directly straight-

Table 3.10 Heckman: comparison FE-results with Heckman selection model, 1995-2006

	Overall		1510		7120	
	FE	Heckman	FE	Heckman	FE	Heckman
1995	-1.705***	-3.199***	-5.453***	– 1.971	-1.314*	– 0.296
	(0.020)	(0.16)	(0.62)	(1.43)	(0.73)	(0.87)
1996	-1.695***	-0.825***	-5.602***	-3.154**	-2.167***	– 1.019
	(0.020)	(0.082)	(0.82)	(1.54)	(0.64)	(0.69)
1997	-1.689***	-0.402***	-6.350***	– 1.629	-1.524***	– 0.575
	(0.021)	(0.072)	(0.85)	(1.03)	(0.48)	(0.72)
1998	-1.708***	-0.268***	-5.740***	– 1.221	-1.457***	-1.463*
	(0.020)	(0.068)	(0.71)	(1.11)	(0.49)	(0.78)
1999	-1.668***	-0.240***	-5.811***	– 0.0825	-1.664***	– 0.733
	(0.020)	(0.067)	(0.58)	(1.19)	(0.45)	(1.37)
2000	-1.764***	-1.606***	-3.230***	– 0.532	-0.868***	– 1.562
	(0.020)	(0.089)	(0.61)	(1.39)	(0.14)	(1.32)
2001	-1.808***	-1.266***	-5.898***	1.823	-1.610***	-9.098***
	(0.020)	(0.059)	(1.09)	(1.15)	(0.59)	(1.99)
2002	-1.892***	-1.339***	-8.474***	– 0.743	-2.107***	-1.464*
	(0.021)	(0.062)	(0.93)	(1.39)	(0.39)	(0.89)
2003	-1.949***	-2.395***	-10.23***	– 0.744	-1.495***	-3.032***
	(0.022)	(0.074)	(0.94)	(1.74)	(0.28)	(0.89)
2004	-1.959***	-2.226***	-6.882***	0.0940	-2.021***	-3.631***
	(0.021)	(0.079)	(0.74)	(1.15)	(0.35)	(0.91)
2005	-1.947***	-2.257***	-4.906***	-2.292***	-1.743***	-2.080***
	(0.021)	(0.075)	(1.10)	(0.84)	(0.34)	(0.56)
2006	-2.985***	-2.050***	-10.12***	– 4.173	0	-3.065***
	(0.11)	(0.087)	(1.26)	(2.62)	(0)	(0.99)
Observations	237528	213582	1666	1636	284	245
Number of censored observations		22578		224		17
rho		-0.886		-0.934		-0.870
athrho (z-statistic)		-183.51		-15.27		-2.98
LR-test of indep.eqns (rho=0): Prob>chi2		0.000		0.000		0.000

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

forward. On the one hand, one can argue that there should be a direct relationship between loss making and competition, because fiercer competition may reduce the profits of the most inefficient firms below zero, and even force them to leave the market. Ignoring those loss making firms when measuring PE can then be justified, since the impact is already reflected in the profits of the (remaining) firms in the sample we use. Hence, the estimation of PE is not biased if loss making firms are ignored.

On the other hand, it can be argued that other issues could affect the level of profits but those are not sufficiently controlled for in our basic model. Those issues could be related to delineating the relevant market including different level playing field (for instance differences in taxes) and product differentiation, or simply bad luck (for instance a fire). This could create a bias in estimating PE.

Using the Heckman selection approach we check for the presence of selectivity bias in our data set (see *e.g.*, Amemiya (1984) and Verbeek and Nijman (1992)).

The results for this approach are reported in table 3.10 for the aggregate level, and for the industries 1510 and 7120. Although the coefficients are all significant for the Heckman results at the aggregate level, they hardly correspond to the FE-results. The results for the two industries differ even more between the two estimation techniques. Notice also, that the coefficients based on the Heckman selection approach are not significant for a number of years, fluctuate widely and sometimes have the wrong sign.

Nonetheless, for every case the test statistics point to selectivity. The correlation (=ρ) between the selection equation (*i.e.* equation that determines making losses) and the basic equation is high. Moreover, the null hypothesis (ρ=0) is rejected if we take the asymptotic t-value as indicator (*i.e.* athrho). An alternative test involves a comparison of the log likelihood for this model with that for a restricted version where ρ=0. Also, this result points to a rejection of the null hypothesis.

Analyzing this selection approach in depth is beyond the scope of this chapter. We leave it for future research to establish whether it is necessary to extend the baseline specification with an extra selection equation. But, the Heckman results for our cases appear to be less convincing than the FE-results. Moreover, compared to the original example that has put the selection bias in the forefront of the literature (*i.e.* whether or not to participate on the labor market), our case is different. Firms (including loss making ones) are already active on the product market. Further, one of the caveats of the Heckman selection model is the choice of the selection variable (*i.e.* which and why?), particularly when data availability is limited.

3.5 Conclusions

This chapter analyzes the robustness of the estimation results of the profit elasticity (PE) using the fixed effect (FE) estimation technique. The chapter provides a guide for researchers how to measure PE in practice and what issues one should be aware of analyzing competition.

In this respect, we advise to estimate PE with a loglinear econometric specification relating the profits and efficiency of a firm. The advantage of such specification is that the slope of the parameter, being our competition indicator PE, is a constant elasticity and therefore convenient in terms of economic interpretation: the percentage increase in profits due to a 1% increase in efficiency captured by average variable costs (AVC). A higher PE indicates more intense competition. Moreover, this specification can be best estimated by using FE as this estimation technique has a number of advantages. The most important one is that it solves the non observed heterogeneity among firms (*e.g.* due to differences in applied technology or management quality) in our panel data set that could be relevant.

Next, the chapter assesses what the effect on the estimated PE is under a range of other conditions to find out whether PE is a robust measure for analyzing the developments of competition over time. These conditions include alternative model specifications, different econometric estimation techniques, and the impact of measurement errors and selectivity issues. For doing so, we employed a data set containing more than 320,000 observations over the period 1993-2006 based on information of about 121,000 individual firms in the Netherlands from 154 industries at the 3-digit SIC-level. We, first, test for the functional form as the functional form of the relationship between profits and AVC is not settled yet. It turns out that the true specification of the preferred model is likely to be loglinear.

Whether one estimation method is more or less appropriate than the other method is basically related to the assumption for the firm specific effect in relationship with the error term of the equation one has in mind. Therefore, we compared the results based on FE, with the results for pooled OLS, random effect, and first difference estimation procedures.

The main message of this econometric analysis is that the idea of the relationship between profits and AVC is to a large extent robust to different ways in which PE can be estimated.

The results of PE based on FE are significantly correlated with the results of pooled OLS, random effect, and first difference estimation procedures. Nonetheless, our preference for the FE-estimation technique using a fixed firm specific effect is supported by the findings of the F-tests and Hausman test (Hausman (1978)) whether the firm-specific effect is fixed or random.

Furthermore, we explore a couple of sensitivity tests to assess the robustness of our FE-model when taking into account potential measurement and selectivity issues in the panel data set. These tests also underline the relevance of this relationship. In terms of correlations, the results for PE are robust compared to other variants that consider measurement errors and selectivity issues, as the correlations are highly significant.

All in all, we conclude that: (i) PE as competition indicator can be estimated in practice using a loglinear specification relating firm's profits to its average variable costs, (ii) the results of PE are to a large extent robust in terms of different econometric estimation techniques, and the impact of measurement errors and selectivity issues.

4 Competition and innovation: Pushing productivity up or down?

4.1 Introduction

This chapter examines the relationship between competition, innovation and productivity at the industry level for the Netherlands.[94] In the view of the endogenous growth theory (see *e.g.*, Romer (1990); Aghion and Howitt (1992) and Aghion and Howitt (2006)), competition and innovation are interrelated and as such seen as important determinants for productivity and in that way contributing to sustained economic growth. And economic growth is a fundamental driver of improving the living standards of the population (*i.e.* the welfare level).

For the Netherlands, this relationship is especially interesting since for many years its performance on productivity growth is relatively poor in an international (and historical) perspective, particularly compared to the US, pushing the Netherlands back in its top-ranking with regard to the level of productivity (see e.g., Van der Wiel (2001a); Gelauff et al. (2004); Kegels et al. (2008) and Van der Wiel et al. (2008)). In that respect, it is not surprising that Dutch policy intends to foster productivity by using policy measures that aim to stimulate either innovation or competition to realize higher welfare.

However, as Nickell (1996) already mentioned more than ten years ago, the theoretical and empirical evidence that competition improves the productivity performance are not overwhelming. Moreover, the study by Aghion et al. (2005) for the UK finds that the relationship between competition and innovation is an inverted U. In that case, a trade-off between both drivers of productivity may exist and innovation policy and competition policy can be at odds with each other when focussed on realizing higher productivity: stimulating competition beyond a certain level might then have a negative effect on innovation, and subsequently on productivity. But empirical evidence is scarce. State of the art research on the empirical relation between competition and innovation as Aghion et al. (2005) did for the UK has not been done for Dutch industries yet, let alone the impact on productivity. Hence, we do not know whether there exists an inverted U curve for the Netherlands, and if so, which industries have

[94] This chapter is based on Brouwer and Van der Wiel (2010). We thank Jan Boone, Roel van Elk, Bert Minne and Ruud Okker for their comments on an earlier version of this paper. We are also grateful to Rob Alessie for his excellent econometric advice.

competition intensities beyond the innovation maximizing level.

This chapter picks up this ambiguous connection between competition and innovation, and relates it to Dutch productivity. We apply an empirical framework that is comparable with Nickell (1996); Griffith et al. (2004) and Griffith et al. (2006). We start from the idea of a production function taking on board views from the endogenous growth theory. Hence, our framework includes the impact competition and innovation have on productivity performance. In the vein of the convergence literature, the distance to the global frontier – as being the highest attainable productivity – level may also be relevant because this may signal potentials for productivity growth through (costless) technology transfers or knowledge spillovers. Recent studies from, for instance, Griffith et al. (2004) and Conway et al. (2006) emphasize the importance of technology transfers and the effect of product market regulations on the international diffusion of productivity shocks given the distance to the frontier. Moreover, our framework both explains changes in competition and innovation, and it provides insight in how the interaction mechanism between competition and innovation works in practice knowing that they are endogenous.

We use data from two sources. First, we employ industry level data from the *Produktie Statistieken* (PS) for more than 150 3-digit SIC-industries directly based on aggregated Dutch firm level data covering almost the whole Dutch economy over the period 1993-2006.[95] Second, we employ innovation indicators from six consecutive Community Innovation Surveys (CIS) covering the period 1996-2006. Moreover, the industry level data is augmented with information from firm level data on variances. Industry averages are sums ignoring firm heterogeneity within an industry, while this is increasingly seen as important in the endogenous growth literature (see *e.g.*, Bartelsman and Doms (2000) and Aghion and Howitt (2006)). As we have firm level data at our disposal, we add measures of variances between firms to our analysis at the industry level to take heterogeneity of firms into account. An example is the distance to the frontier.

To some extent, the ambiguous message from the empirical literature on the relationship be-

[95] For a number of industries, we use 4-digit SIC-industries.

tween competition and innovation is related to the difficulty in measuring both concepts and the availability of adequate data. Moreover, from a modeling perspective, both competition and innovation are endogenous and this complicates estimation of their impact on productivity. Other factors may determine competition and innovation like policy measures and macro economic shocks. Additionally, competition may affect innovation as well, but innovation may also change the market structure and the degree of competition through product differentiation. Hence, we have reverse causality and encounter endogeneity problems (see also chapter 5). In the current chapter, we address these problems using Generalized Methods of Moments (GMM). This estimation technique exploits lagged explanatory variables as instruments to cope with endogeneity.

Our main findings are the following. We show that competition and to a lesser extent innovation are good for productivity. But here it is important to take into account the relationship between competition and innovation. We provide support for the view that there exists an inverted U-curve between competition and innovation for the Netherlands, at least for the manufacturing sector. This corresponds with findings of Aghion et al. (2005). As there can be a trade-off between both, our findings have implications for policy because competition and innovation might be in conflict. However, we show that the overall results indicate that a negative effect of competition on productivity through lower innovation arises only at very high levels of competition. Hence when it comes to productivity, more intense competition is almost always better.[96]

This study contributes to the (empirical) literature in different ways. First, it examines the existence of an inverted-U curve between competition and innovation for the Netherlands. Besides the study of Aghion et al. (2005), empirical evidence for such inverted U-curve is scarce.[97] Second, compared to Aghion et al. (2005), we use better measures for competition and innovation. Aghion et al. (2005) applies the price cost margin (PCM) and patent citation as indicators for competition and innovation respectively. Both indicators have severe shortcomings. We use the profit elasticity (PE) and the innovation rate as those indicators are

[96] This is in line with findings in chapter 5.

[97] See Creusen et al. (2006b) for an analysis of the inverted-U curve for only the Dutch retail trade.

more robust (see *e.g.*, Boone et al. (2007a), Kleinknecht et al. (2002) and Brouwer (2007)). Moreover, our study analyzes the entire economy, whereas Aghion et al. (2005) only look at manufacturing industries. We also have data for other industries like services. Third, we explicitly consider the effect of competition and innovation on productivity, since the latter is one of the main goals for policy makers as productivity is directly related to welfare. As far as we are aware of, there is no comparable study in this respect. Finally, we control for possible feedback mechanism from innovation to competition taking account of other explanatory variables.

The structure of this chapter is as follows. Section 4.2 gives a brief theoretical background of the relationship between competition, innovation and productivity. The empirical framework, econometric specification and endogeneity problems are discussed in section 4.3. Next, section 4.4 introduces the data sources and the key variables in question. It also presents some descriptive statistics. Section 4.5 contains the results of estimating the relationship between competition, innovation and productivity. This section also examines the robustness of these results with respect to our competition indicator. Finally, section 4.6 summarizes the main findings and sketches policy implications.

4.2 Theoretical and empirical background

4.2.1 Theoretical background competition and innovation

Traditional views competition and innovation

Let us first have a closer look at the separate effects of competition and innovation on productivity. In theory, investments in R&D create new technologies and new products, both generating higher productivity, or stated otherwise: more value added per worker (see *e.g.*, Grossman and Helpman (1991); Cameron (1998); Griliches (1998) and Cameron and Trivedi (2005)). The general finding in empirics is that innovation is good for productivity (see, inter alia, Van Leeuwen (2009)).

The intensity of competition is also important for economic growth (see *e.g.*, Geroski (1990) and Nickell (1996)). This can be found in theory and empirics.[98] The story goes

[98] Exceptions are Scharfstein (1988) and Martin (1993), they argue that competition leads to an increase in

as follows. Competition on product markets is generally seen as generating lower prices for consumers. Competitive pressure stimulates firms to operate efficiently by, for instance, *cutting the fat out* of their organizations. Or, more intense competition forces inefficient firms to leave the market. It brings product prices in line with their marginal costs, lowering the rents of producers and increasing consumer surplus. Vigorous product market competition may therefore result in higher productivity as resources and output are allocated to their most productive use.

However, taking into account the interplay between (product market) competition and innovation, economic theory does not predict the shape of this relationship nor how competition affects productivity and economic growth through innovation. Whether or not competition raises innovation is an ongoing debate and a challenging research topic since Schumpeter's remarks in two famous books, dividing the theoretical strands into two camps. The first strand consists of those that argue that competition is bad for innovation (see Schumpeter (1942)). The second strand claims that competition is good for innovation (see Schumpeter (1934)).

According to Schumpeter (1942) fiercer competition generates less R&D, reducing the rate of innovation and hence economic growth. The intuition is that because the expectation of high profits drives innovation, an increase in competition will discourage innovation if it results in lower profits. The Industrial Organization literature of product differentiation and monopolistic competition supports this strand (see Salop (1979) and Dixit and Stiglitz (1977)). Using a Schumpeterian endogenous growth model, Aghion and Howitt (1992) show that an increase in product market competition has a negative effect on productivity growth by reducing the monopoly rents that reward innovation (see also Romer (1990) and Grossman and Helpman (1991)). Examples of empirical studies that support this negative correlation are Hamberg (1964); Mansfield (1964); Kraft (1989); Porter (1990) and Symeonidis (2001).

The view that competition is good for innovation, is theoretically supported by studies from Schumpeter (1934); Arrow (1962) and Scherer (1980). In this strand, it is thought that competition stimulates an incumbent to innovate otherwise the firm is forced to leave the market and the potential entrant will win the race. This entrant will win this race if the replace-

managerial slack, and hence lowers productivity.

ment effect (Arrow (1962)) for the incumbent is stronger than its efficiency effect.[99] When innovating the incumbent monopolist replaces her own profits while the potential entrant has no pre profits to replace at all. Aghion and Howitt (1999) show these mechanisms in a theoretical model. More intense competition raises innovation activities, because it reduces incumbent's pre-innovation profits more than it lowers its post innovation profits. The empirical evidence for this second strand is larger than for the first strand. We refer to studies like Geroski (1990); Nickell (1996), Blundell et al. (1995, 1999) and Carlin et al. (2004) that find a positive relationship between competition and innovation (or productivity).

Recent view: nonlinear relation competition and innovation

Having both a positive and a negative relationship in the literature, the third strand in the debate is predictable: the connection between competition and innovation can be characterized as an inverted U-shape. Reconciling theory and empirical evidence, Aghion et al. (2005) develop a model where low (high) levels of competition have a positive (negative) effect on innovation.[100]

The intuition behind this inverted-U is as follows. There are two types of competition effects on innovation: escape competition effect and Schumpeterian effect.

In case of low levels of competition, the escape competition effect dominates. While preinnovation profits are reduced, increasing competition will raise the incentive of neck-and-neck firms to innovate because firms become the single front technology if they innovate. But if competition further intensifies, the balance between the two effects changes and the Schumpeterian effect (*i.e.* fiercer competition generates less R&D) will start to dominate, generating the negative part of the inverted-U curve between competition and innovation. Further increases in competition reduce the (post)innovation rents for laggard firms to become neck-and-neck with the leader again.

Hence, the inverted-U relationship arises due to a change in the composition of firms. Initially when competition is low, industries are most often leveled. So if competition increases industries become more frequently unleveled, whereas the chance that they become

[99] When the monopolist does not innovate, he loses his current monopoly profits. This gives an incentive for the monopolist to innovate.

[100] Note that it was Scherer (1967) who for the first time came up with the idea of an inverted-U relationship.

leveled again reduces as for laggards it is increasingly difficult and costly to catch up. Stated otherwise, when competition is really fierce hardly any industry will remain leveled. Consequently, as the innovation rate is lower in unleveled situations, beyond some level of competition, innovation will decline, generating the inverted U. Moreover, Aghion et al. (2005) add the idea of neck-and-neck industries (or firms) where the difference in performance is small across firms as they have the same technology, whereas in 'leader-follower' industries firms have different technologies and, hence, different productivity levels. Due to more neck-and-neckness, the inverted U becomes steeper as the escape competition is larger.

Such inverted U-curve between competition and innovation can also occur in another way as there can be a trade-off between process and product innovation as well when competition is raised (see Boone (2000b)). At the industry level, this may then generate an inverted U-curve if total innovation expenditures (*i.e.* process and product outlays) are considered.

Boone (2000b) shows that a rise in competition may raise industrywide efficiency through more process innovation. But, this may reduce product variety or the number of products introduced to the market: less product innovation. The reason is that when competition becomes more intense, inefficient firms are forced to leave the market as a result of the selection effect of competition and lower costs of opponents (higher efficiency level from process innovation). This reduces the product variety (or product innovations) in this market. Moreover, more competition reduces profits and makes it for some firms less attractive to introduce a new product. Hence, a trade off may occur between process and product innovations at the aggregate level.

There are, however, two possibilities that may overturn this trade off. First, firms could also escape competition by product differentiation, and hence creating their own niches (see also chapter 5). Second, lower profits due to more competitive pressure could act as a wake up call for managers. To avoid bankruptcy, managers have to look for new products that can generate additional profits. Hence, although process innovation is applied industrywide, innovation expenditures with respect to product innovation might go up as well in that particular industry.

The empirical evidence for an inverted U-shape between competition and innovation is scarce.[101] Besides Aghion et al. (2005), only Scott (1984) and Kilponen and Santavirta (2007) found significant evidence in favor of this form. For instance, the latter examines this relationship for the manufacturing companies in Finland between the years 1990 and 2001. In contrast, tested on a data set of Swedish firms, Tingvall and Poldahl (2006) find that the inverted-U relationship relation between competition and R&D is by PCM. Further, using firm level data Creusen et al. (2006b) tested the inverted U-shape for the Dutch retail trade but their results also rejected this view. Finally, Griffith et al. (2006) find no evidence for an inverted U-curve looking at an unbalanced panel of nine countries and 12 two-digit manufacturing industries over the period 1987-2000.

The following summarizes this section. Positive as wells as negative effects from competition on innovation can be found in theory and empirics, while recent literature suggests an inverse U relation between competition and innovation. Consequently, the implications for productivity are similar: effects can be positive and negative. But recent theory indirectly provides indications that fiercer competition is always good for productivity. As a matter of fact, from Aghion et al. (2006) one can deduct that a decline in innovation expenditures (of incumbents) in an industry can go hand in hand with higher aggregate productivity of that particular industry. The reason is that after intensifying competition, the least efficient domestic firm has no incentive anymore to imitate or to innovate due to the large productivity gap to the technological frontier. So the innovation expenditures of that industry decline. Yet it can be proved that aggregate productivity of that industry rises (see Kocsis et al. (2009)). The reason is the entry of a foreign leader with the highest productivity level in that particular industry. That foreign firm replaces the least efficient domestic firm increasing aggregate productivity, but this entry is not seen (in statistics) as an innovation.

4.2.2 Further extension endogenous growth literature: distance to frontier

Following up on the fundamentals of earlier work of Aghion and Howitt (see Aghion and Howitt (1992) and Aghion and Howitt (1999)), the idea has been postulated that the distance

[101] Below in subsection 4.2.3, we show that this finding might also depend on the indicators used for competition and innovation.

to the technological frontier (*i.e.* the technology giving the highest possible level of output given the inputs) matters for countries or industries. For example, Aghion and Howitt (2006) describe a model where the growth performance of a country (or industry) also depends on its proximity to the technological frontier and what both innovation and competition mean in this respect.

A similar story can be told for the interaction between entry and the distance to the frontier. In this model (see Aghion et al. (2006)), entry threat is an exogenous parameter which measures the probability that a (foreign) firm enters the (home) market. The results of this model are the following. The impact of entry on innovation is non-uniform across firms and industries. Higher threat of entry leads to higher innovation expenditures and higher productivity growth of incumbents, which are already highly efficient (*i.e.* firms close to the frontier). These firms innovate more to prevent entry. However, increased entry (threat) discourages less efficient incumbents (*i.e.* firms far below the frontier) to spend on innovation. The reason for the heterogeneity in the incumbents' response to the entry threat is simple: while the costs of innovation are the same for all firms, the market leaders have a higher chance of retaining their leadership in the face of entry than the laggards have of gaining it.

4.2.3 Empirical issues

Linking competition, innovation and productivity is not only from a theoretical perspective an unsettled issue. Dealing with it in practice is a challenging case too and still in its infancy. The ambiguous empirical results with regard to competition and innovation may partly be related to doubtful indicators for competition and innovation. Two questions are highly ranked on the research agenda: (i) How to measure competition?, and (ii) how to measure innovation?

In the empirical literature, competition is often measured with variables like concentration, profitability, price cost margins (see *e.g.*, Domowitz et al. (1986), Blundell et al. (1995, 1999), Nickell (1996), and Aghion et al. (2005)). Boone (2000a) and Boone et al. (2007a) have shown that these competition measures are not monotone in competition. When competition intensifies due to more aggressive interaction between firms, the industry PCM may rise suggesting less competition. The reason is that PCM increases as a result of the reallocation of market shares from inefficient firms (with low mark ups) to efficient firms (with high mark ups). This chapter uses, therefore, the profit elasticity (PE) as indicator for competition. This measure relates the firm's profit to its efficiency that can be captured by average variable costs. The intuition behind this indicator is that inefficient firms are punished more severely in terms of their profits when competition intensifies. This measure of competition is monotone for different parameterizations of competition (see *e.g.*, Boone et al. (2007a) and Boone et al. (2010a)).

This chapter employs the innovation intensity – innovation expenditures over employees – as indicator for the innovation activities instead of, for instance, R&D expenditures or patents. We do not use the R&D measure as this measure does not cover all the innovative efforts of firms.[102] The definition of innovation expenditures we use, is much wider than the one for R&D that is often used in the studies mentioned above. Our innovation indicator consists of, amongst others, costs of patent application, wages of R&D personnel, exploitation costs, and capital expenditure on buildings and equipment for R&D.

The same limitations pertains to the number of (applied for) patents or cited patents as indicator for innovation. This indicator is for example used in Aghion et al. (2005) and Kilponen and Santavirta (2007). The problem, here, is that not every innovative firm applies for a patent due to, amongst others, high costs of application and the desire to keep the innovation secret. This shortcoming is particularly relevant in non-manufacturing industries that we want to analyze as well. Innovations in (particular) services can hardly be patent. Hence, these industries would then be excluded from further analysis if an innovation measure based on patents is used.[103]

Finally, the studies from Tingvall and Poldahl (2006) and Aghion et al. (2005) illustrate that the results of the inverted U-shaped relation between competition and innovation can be sensitive to either the choice of competition measures or the innovation indicator. Tingvall and Poldahl (2006) find strong support for the inverted-U relationship using the Herfinfahl index (H). However, if this concentration indicator is replaced by PCM, then they do not find support for this form. Similarly, Aghion et al. (2005) do not find a statistically significant inverted U-shape when they use R&D-expenditures as indicator for innovation. Below we test the sensitivity of our results for using different competition measures.

[102] In our data set, the R&D expenditures are not even half as much as the innovation expenditures.

[103] For instance, Kilponen and Santavirta (2007) excluded the industries without any US patents.

4.3 Econometric specifications

4.3.1 Empirical framework

The basic idea in our framework is that both competition and innovation are major determinants of productivity, and productivity is one of the main goals for policy as productivity is the direct link to welfare (see also chapter 1).[104]

Our empirical model consists of components of studies from Nickell (1996); Griffith et al. (2004) and Griffith et al. (2006). It integrates the views of existing literature such as the two faces of R&D, the convergence debate and the existence of firm level heterogeneity in productivity.

We start with a production function taking on board mechanisms from endogenous growth theory that both innovation and competition matter for economic growth. Therefore, in contrast to Nickell (1996) who focuses on the impact of competition on productivity, our model also includes the impact of innovation on productivity performance. As Griffith et al. (2004), we take into account views from the convergence literature and the role of the so called two faces of R&D (see also Cohen and Levinthal (1989)), where convergence between countries/industries depends on the absorption capacity of knowledge spillovers.[105] Finally, as competition and innovation are both endogenous variables, we explain these variables separately in our model.[106]

Productivity equation

Assume that each industry j produces in period t according to a standard neoclassical production technology

$$Y_{jt} = AF(L,K) = A_{jt}K_{jt}^{\alpha}L_{it}^{\beta} \tag{4.1}$$

[104] Similar to innovation and competition, human capital (or human skills) may have an impact on productivity. Seen as another input factor in the production process, human capital might help to speed up technology absorption and stimulating innovation. Sianesi and Van Reenen (2003) provide a comprehensive overview of empirical studies on the effects of human capital on growth. However, in this chpater, we ignore human capital as driver of productivity, because we have no data on human capital.

[105] Notice that Griffith et al. (2004) neglect competition issues.

[106] Competition determines innovation. But there may also be reverse causality as innovation may affect competition (see also chapter 5).

where Y is (real) output, K denotes capital, L is labor, and A indicates total factor productivity (TFP). We assume that the elasticities of capital and labor (*i.e.* α and β) exhibit diminishing marginal returns to the accumulation of each factor alone and these elasticities are constant over time and across industries.[107] A is allowed to increase over time. Taking the natural logarithm, we write equation (4.1) as a decomposition of labor productivity (LP) growth into contributions of the capital intensity, the shifts in the industry's size (in terms of employed staff) and A[108]

$$\Delta lp_{jt} \equiv \Delta y_{jt} - \Delta l_{jt} = \Delta a_{jt} + \alpha\left(\Delta k_{jt} - \Delta l_{jt}\right) + \left(\alpha + \beta - 1\right)\Delta l_{jt} \tag{4.2}$$

Note that the parameter on industry size (*i.e.* labor) determines whether the firms in industry j can benefit from increasing economies of scale (*i.e.* if $\alpha + \beta - 1 > 0$).

The view of the endogenous growth theory that innovation and competition matter for growth enters our equation through A (see *e.g.*, Romer (1990); Griliches (1998); Aghion and Howitt (1992); Nickell (1996); Griffith et al. (2004) and Aghion et al. (2006)).[109] Taking a closer look at the determinants of A (or TFP), we assume that industries may enhance their productivity growth in four ways.

First, based on theory, we expect that fierce competition forces firms in a particular industry to reduce 'X inefficiencies', and consequently affects productivity in the short term (see for instance Nickell (1996), for an overview). Weak competition makes managers and employees lax, or even seduces managers and employees to shirk.

Second, based on the convergence literature, the Schumpeterian growth theory takes into account the distance to the technological frontier as a measure of the potential for technology transfer. The larger the distance is the further firms lie behind the frontier and the greater the potential of productivity growth through technology transfers within an industry. For instance, Griffith et al. (2004); Conway et al. (2006) and Van der Wiel et al. (2008) empirically

[107] That is: $\alpha, \beta \in <0, 1>$.

[108] Lower case letters mean logarithm of the variables concerned.

[109] In the endogenous growth literature, there is an ongoing debate about semi- versus endogenous growth theory (*i.e.* Schumpeterian growth theory). Roughly speaking, according to the semi endogenous growth models one should estimate a productivity equation in levels, whereas according to the endogenous growth theory one should estimate growth rates. See Madsen (2008) for a further discussion and why (time-series) evidence is more favorable for the Schumpeterian growth theory using growth rates.

show that the distance to the frontier matters for productivity growth. We examine whether or not this 'gap'(=g) also affects growth rates of TFP.

Third, innovation might have a direct impact on the rate of TFP growth by conducting R&D to develop new process technologies and/or new products (so called first face of R&D). Although, knowledge has the characteristics of a public good (knowledge spillovers), changes in TFP require real resources in terms of R&D (and human capital) to exploit those knowledge spillovers, but also to generate knowledge in the first place.

Finally, it can be argued that the ability of a firm (or industry) to benefit from knowledge spillovers depends on its own level of R&D activities and the distance behind the technological frontier. This idea is developed by Cohen and Levinthal (1989), who established the concept of the 'two faces of R&D'. In fact, R&D activities play two roles. On the one hand, R&D activities generate innovations. On the other hand, R&D improves the ability of a firm to identify, assimilate and exploit outside knowledge. Cohen and Levinthal (1989) label this as the learning or absorptive capacity of the firm. The absorptive capacity is largely a function of the firm's level of prior knowledge (see also Griffith et al. (2004)).

Going back to our model and include the preceding elements, we first assume that A depends on the stock of knowledge (S) and the intensity of competition (C). Industries with a larger (R&D) knowledge stock or more intense competition have a higher level of TFP. Taking logarithms and differencing with regard to time, the rate of A depends on the growth rate of S and the change in C and X

$$\Delta a_{jt} = v_0 \Delta s_{jt} + v_1 \Delta C_{jt-1} + v_2 \Delta X_{jt-1} \tag{4.3}$$

with $v_0 > 0$, $v_1 > 0$ and X is a vector of control variables which (in theory) may include other (exogenous) explanatory variables that affect TFP-growth like non-technological innovations or spillovers from outside the industry. We assume that competition and those other determinants do not directly affect TFP but with a lag. In doing so, we also eliminate already some of the endogeneity bias in our framework.

Combining equations (4.3) and (4.5, see box) gives an expression for TFP which depends on competition as well as innovation, and on a vector of control variables

$$\Delta a_{jt} = \mu_1 IR_{jt-1} + v_1 \Delta C_{jt-1} + v_2 \Delta X_{jt-1} \tag{4.6}$$

Intermezzo: Logs or no logs for innovation intensity?

On the one hand, one can assume that the stock of industry knowledge only increases with the industry's innovative effort (=I)

$$S_{jt} = \delta S_{jt-1} + I_{jt-1} \tag{4.4}$$

Assuming that depreciation of knowledge stock of capital is absent ($\delta = 1$) and using

$$\nu_0 \Delta s_{jt} = \nu_0 \frac{I_{jt-1}}{S_{jt-1}} = \mu_1 \frac{I_{jt-1}}{Y_{jt-1}} = \mu_1 IR_{jt-1} \tag{4.5}$$

Where $\mu_1 = \frac{\partial Y}{\partial S}$, $\nu_0 = \frac{\partial Y}{\partial S}\frac{S}{Y}$, and IR innovation intensity. Then the capital knowledge stock in equation (4.3) can be replaced for the innovation intensity.

On the other hand, it is also defendable to use logs (I) in equation (4.3) instead of the innovation intensity because IR can be part of the production function in a neoclassical framework in the same way as the other input factors K and L. Note that then the meaning of a change in A becomes different as it does not include changes caused by innovation anymore. Innovation is treated as a separate input variable. We use the former way.

Implementing equation (4.6) in the productivity equation (4.2), we obtain:

$$\Delta lp_{jt} = \mu_1 IR_{jt-1} + \nu_1 \Delta C_{jt-1} + \nu_2 \Delta X_{jt-1} + \alpha\left(\Delta k_{jt} - \Delta l_{jt}\right) + \left(\alpha + \beta - 1\right)\Delta l_{jt} \tag{4.7}$$

Finally, we include the distance to the frontier and the second face of R&D to this equation. Summarizing we have:

$$\begin{aligned}\Delta lp_{jt} = {} & \mu_1 IR_{jt-1} + \nu_1 \Delta C_{jt-1} + \delta_1 g_{jt-1} + \delta_2 g_{jt-1} IR_{jt-1} + \delta_3 g_{jt-1} \Delta C_{jt-1} + \nu_2 \Delta X_{jt-1} \\ & + \alpha\left(\Delta k_{jt} - \Delta l_{jt}\right) + \left(\alpha + \beta - 1\right)\Delta l_{jt} + T_t + \varepsilon_{jt}\end{aligned} \tag{4.8}$$

where $g = ln(A_f/A)$. This term captures the gap and in that sense the potential technology transfers from the technological frontier (A_f).[110] To estimate this equation, we add an error term (ε_{jt}) to this equation assuming that this is serially uncorrelated. Moreover, we include time dummies T to control for macroeconomics shocks that affect TFP in all industries.

[110] The pace of this catch up depends on the size of the estimated coefficient δ. For instance, patents may hamper spillovers and lower this coefficient.

Innovation equation

What determines innovation? According to Tingvall and Poldahl (2006), there is no explicit theoretical model with preferred explanatory variables. From section 4.2 we infer that competition is important. The degree of competitive pressure affects the amount of investment in innovation. Following Aghion et al. (2005) the nonlinear relation between competition and innovative effort can be estimated by regressing the innovation rate of each industry on a quadratic function of competition intensity in the respective industry. Then the equation for the innovation rate for industry j in period t is

$$IR_{jt} = \varphi_1 C_{jt-1} + \varphi_2 C^2_{jt-1} + \varphi_3 W_{jt-1} + T_t + \psi_{jt} \tag{4.9}$$

with W being other determinants of innovation like policy measures in the form of subsidies and the possibilities of cooperation between firms. We use lags as we assume that the impact of our explanatory variables takes some time to affect innovation. We also add an error term (ψ) and time dummies to this equation.

Theory provides some guidance for the parameters φ_1 and φ_2 as discussed in section 4.2. If $\varphi_1 > 0$ then for ΔC close to zero, the dominant effect is escape competition: firms innovate more when competition intensifies. In contrast, $\varphi_1 < 0$ then for ΔC close to zero, the Schumpeter effect dominates the effect of competition on innovation. In this case, competition discourages the innovative efforts of an industry as (laggard) firms find it difficult to reap the benefits of these efforts. But following Aghion et al. (2005) this relationship may also be nonlinear with $\varphi_1 > 0$ and $\varphi_2 < 0$.[111] If competition is low then competition is conducive to I, whereas if competition is high then competition may discourage I.

Competition equation

Given the complexity of modeling competition, our aim here is to estimate a simple equation relating competition to a number of determinants at the industry level.[112] We model competition as follows

$$C_{jt} = \lambda_1 IR_{jt-1} + \lambda_2 IR^2_{jt-1} + \lambda_3 Z_{jt-1} + T_t + \zeta_{jt} \tag{4.10}$$

[111] The downward sloping part of the inverted U-shape occurs beyond the level of C where: $2\varphi_2 C > \varphi_1$.

[112] As far as we know, empirical research that may serve as a reference is scarce (see Creusen et al. (2006a) for one of the exceptions).

with Z a vector of other explanatory variables discussed below. Theory has put forward several (exogenous) determinants of competition (see *e.g.*, Tirole (1988); Cabral (2000) and Boone (2000a)). Some of these determinants are related to market structure of industries and conduct of firms. Given our available data, Z includes variables that are linked to strategic entry barriers such as advertising costs and number of firms that enter and exit the market.[113]

We put I into equation (4.10) to take account of a possible feedback mechanism from innovation back to the intensity of competition at the industry level. The idea is that the higher the competition intensity in an industry, the higher the incentive for firms to reduce the competition intensity by differentiating their products from that of their competitors by creating niches. Hence, when the outlays for product innovation increase, this may eventually have a negative effect on the degree of competition. We use a one year lag as to take into account that our innovation indicator is an input measure and does not directly affect the extent of competition.

4.3.2 Industry averages and heterogeneity of firms

As explained below, this chapter uses industry level data to limit the complexity of the econometric model. These industry data are averages based sums from firm level data.

However, we do not completely ignore information based on firm level data because it is well known that firms are heterogenous in their innovative efforts (see *e.g.*, Bartelsman and Doms (2000); Van der Wiel and Van Leeuwen (2003); Bartelsman et al. (2004) and Van Leeuwen (2009)). This chapter therefore links firm level data to industry level data to take account of the possibility of different responses of firms instead of assuming a representative firm response within an industry.[114] First, we already discussed the importance of the distance to the frontier as driver of industry's productivity growth. Having firm level data at our disposal, we can consider the relevance of this issue. Second, we also control for variances in efficiency per industry. More precisely, we add the variance of the average variable costs as control variable to the innovation equation (4.9) and to the productivity equation (4.8).

[113] Notice that these variables are not exogenous themselves.

[114] This is an interesting field for further research. As firm level data is most often confidential, statistical offices could add moments of variables based on firm level data without violating confidentiality. This opens a new dimension for research at the industry level.

The reasoning is that, at high levels of competitions, firms will adopt or use the existing technology quicker/better if the variation is small than when it is large (*i.e.* reducing X-inefficiencies).[115] Moreover, we relate this difference in efficiency to the extent to which industries are neck-and-neck. Aghion et al. (2005) argue that when industries are more neck-and-neck (*i.e.* lower variance in efficiency across firms, or stated otherwise, firms operate at similar technological levels) the more positive the effect of competition on innovation. If variance is high, then an increase in competition will have a stronger negative effect on innovation. All in all, the peak of inverted U will be higher and occurs at a higher level of competition.[116]

4.3.3 Econometric issues

One difficulty in analyzing the relationship between competition and innovation is that both factors are not exogenous. In fact, competition might even be endogenous due to reverse causality with innovation. To illustrate, innovation can affect competition in two ways. First, high R&D-investments can reduce entry as if other firms have to follow this they form a barrier to entry thereby reducing competition (see Sutton (1991)). Second, innovation can take the form of product differentiation thereby reducing competition by creating niches and by making goods less perfect substitutes (see Boone (2000b)).

The study of Aghion et al. (2005) uses a set of policy instruments to cope with the endogeneity of competition due to innovation. These instruments (*e.g.*, privatization, EU Single Market Program, and Monopoly and Merger Commission investigations) are based on the introduction of policy changes across industries. These changes are likely exogenous because they are not related to innovation performance. Unfortunately, a similar data set with policy changes is currently not available for the Netherlands. We need another approach.

We use GMM estimation technique to cope with endogeneity problems. GMM exploits lagged explanatory variables as instruments after the equation has been differenced to eliminate unobserved fixed effects. To be more precise, our model consists of the three earlier

[115] We do not include this variable into the competition equation due to probably high collinearity.

[116] Aghion et al. (2005) state that the fraction of sectors with neck-and-neck competitors is itself endogenous, depending upon equilibrium innovation intensities. But, in our view, lower variance could also be the result of intensifying competition selecting the best performing firms from inefficient firms and making the difference between the remaining firms smaller.

mentioned equations: productivity, innovation and competition in equations (4.8), (4.9) and 4.10 respectively. These equations are estimated in first differences and all right hand side variables in our model are lagged with one year. Subsequently, the endogenous variables on the right hand side are instrumented with all the exogenous variables of the model including the second and third lagged of the endogenous variables themselves. Of course every instrument is the same for each endogenous variables on the right side. The first-stage regressions, where we estimate the endogenous variables on the right hand side with the instruments, are tested with the Hansen's J test (test of over identifying restriction) and the GMM C statistic (test of endogeneity).

GMM estimation technique is to be preferred above for instance IV-techniques in the following situations. In case of heteroskedasticity the IV-estimates of the standard errors are inconsistent, and also the tests for endogeneity and overidentifying restrictions are then invalid (see Baum et al. (2003)). When facing heteroskedasticity of unknown form, GMM is the estimation approach. GMM makes use of the orthogonality conditions to allow for efficient estimation in the presence of heteroskedasticity of unknown form (see Hansen (1982)). One of the advantages of GMM is also that it can estimate the coefficients in a model without solving the model analytically (Verbeek (2004)). Therefore we can estimate our three equations separately.

For this analysis, we use industry level data instead of firm level data because we are then able to estimate our complete model with fixed effects regressions and instruments using GMM. We see this analysis as a first step to analyze firm level data in future research. An analysis of that type encounters a number of (econometric) challenges to deal with that we now can circumvent using industry level data. For instance many firms do not innovate at all, because of that a Tobit or Heckman model combined with two fixed effects regression equations for the other two equations is required. Also estimating a fixed effect Tobit is not that easy as extra assumptions are needed. Finally, at the firm level we have many missing observations for innovation because all firms below 50 employees are sampled and there is not much of a chance that a firm is present in the sample for the entire observed period.

4.4 Data description

4.4.1 Data sources

We use a number of data sources. The most important ones are: Produktie Statistieken (PS) and Community Innovation Survey (CIS). Both sources are surveys from Statistics Netherlands and based on firm level data. Below, we briefly describe these two main sources of information in more detail.

PS

Data on, for instance, labor productivity is derived from PS, produced by Statistics Netherlands on a yearly basis. Data from PS is available for the years 1993 to 2006.[117] The PS is a sampled survey; only firms with more than 20 employees are included in the sample each year. For smaller firms, sampling fractions decrease, and consequently most smaller firms will have gaps in the data for several years. Moreover, Statistics Netherlands apply a rotating sample method to reduce the administrative burden of (small) firms. This also reduces consecutive observations of firms.

CIS

Data on innovation expenditures has been gathered from the Dutch section of CIS. CIS is a European harmonized questionnaire, held every two years, containing questions about innovative activities in enterprizes. Our innovation data covers the period 1996-2006. In fact, we use six consecutive CIS-surveys: *i.e.* CIS2 for 1994-1996, CIS2,5 for 1996-1998, CIS3 for 1998-2000, CIS3,5 for 2000-2002, CIS4 for 2002-2004, and CIS2005 for 2004-2006. CIS samples firms below 50 employees. Firms with less than ten employees are not included.

A main advantage of CIS is that after merging with PS one can directly relate firms' innovation activities to their performance and input factors. Yet CIS has shortcomings that limit the options for research. We mention the most important ones. First, the number of observations in CIS is low compared to that of PS due to a more limited sampling technique. This narrows the matching with PS. Additionally, CIS contains industries that are not present in PS and

[117] Data for the industries transport and telecom only covers the period 2000-2006.

vice versa. This reduces the number of industries that can be examined. Second, CIS suffers from lower response rates and the responses can be selective as it is most likely that innovative firms are more inclined to respond than firms that do not innovate. Finally, CIS does not capture all issues of innovation. For example, information on human capital formation is not included.[118] Also, new firms entering the market are initially not included in the sample, while these firms may enter the market because they are innovative.

Taking the caveats of our sources for granted, after aggregating firm level data to industry level data, we merged the two data sources at the 3 (and sometimes 4) digit SIC-code in order to obtain information over the period 1996-2006, and in order to be able to construct lagged exogenous variables that we need for our estimation technique later on. Because we do not have CIS data in odd years, we lose observations. To keep enough observations, we interpolate the innovation data which may reintroduce some endogeneity.[119]

4.4.2 Variables

This subsection discusses the definitions of our dependent variables and the explanatory variables respectively that we use for estimating the equations for productivity, innovation and competition.

Labor productivity

Labor productivity is defined as gross value added per employee, and is derived from PS.

Innovation intensity

The expenditure on innovation divided by the number of employees is used as a measure of the innovation intensity of an industry.[120] As explained in equation (4.5), we use a ratio and this ratio comes from CIS. The innovation expenditures consist of the total costs of both contracted R&D and intramural R&D, including wages, exploitation costs, and capital

[118] Some European countries like Finland do take human capital issues into account.

[119] We use a linear interpolation. Besides having less observations, if we do not interpolate it is hardly possible to use GMM as we need observations for the years t, $t-1$ and $t-2$.

[120] We do not use sales in the denominator because, it turned out that the sales from CIS were not reliable. An alternative not applied here is to use the sales from the PS.

expenditure on buildings and equipment for R&D.[121]

Measures of competition

With the data at hand there are several routes open for measuring competition. In this chapter we use PE, (see Boone et al. (2007a)). This measure results from an econometric specification that relates profits to efficiency captured by the average variable costs. This regression is applied to firms belonging to one and the same market (or industry). The parameter of this regression measures PE and comparing this parameter over time enables us to make inferences on changes in competition. The main idea of PE is that fiercer competition enables efficient firms to earn relatively higher profits than their inefficient competitors. PE measures the percentage fall in a firm's profits in response to a 1 percentage increase in the firm's cost per unit of output.

An alternative measure for the extent of competition is the PCM. This measure refers to the firm's ability to set its prices above its marginal costs. This chapter defines PCM at the industry level as gross profits as a proportion of total sales. Gross profits is value added minus total wages and the costs of intermediate inputs.

Both competition measures are based on firm level data from PS.

Physical capital

Physical capital is an input factor in the production process that determines output (see equation (4.1)). Unfortunately, time series for this type of capital are scarce, particularly at the firm level. Indeed, as we use an unbalanced panel data set based on a sample, it is very hard to construct a capital input measure for each firm in the data set as firms are not present in all consecutive years. Therefore, we employ an alternative indicator at the industry level. We aggregate all the depreciation expenditures within an industry. In fact, we use the depreciation rate (*i.e.* depreciation expenditures over gross value added) as measure for the capital intensity as can be deducted from equation (4.6). Figures originate from PS.

[121] Although the difference between product and process innovation expenditures can be important from a theoretical perspective (see Boone (2000b)), we cannot distinguish between both concepts as CIS does not provide separate figures for either product or process innovation expenditures.

Non-technological innovations

Non-technological innovations in CIS are defined as changes in strategy, management, organization, or marketing. This type of innovation can enhance the performance of a firm or an industry. Particularly, firms may realize higher productivity gains if they simultaneously do technological and non-technological innovations (see Hempell et al. (2004)) than doing either technological or non technological innovation suggesting that those innovations are complementary. Put differently, technological innovations might be a necessary condition for improving the performance of a firm, but not a sufficient condition.

CIS provides only discrete data (yes or no) and no data on outlays for non technological innovations. We employ the percentage of firms (as percentage of total number of firms in an industry) that implement a non-technological innovation.

Distance to the frontier

The distance to the frontier (g) can be a determinant for productivity as explained in section 4.2. Due to data availability, for this study we limit ourselves to data for the Netherlands.[122] In theory, the highest productivity level of all firms in a given Dutch industry represents the (national) frontier. However, defined in such way this definition for the frontier is very sensitive to the presence of outliers in the data. To reduce this sensitivity, we look at the highest quartile in the labor productivity distribution in each 3-digit SIC class instead of the highest single labor productivity level of one particular firm in that industry. The productivity level of these firms in this quartile will be taken as the frontier and this level is related to the average productivity level of the industry to measure the (average) distance to the technological frontier. We expect positive estimated coefficients for g, including the interaction terms that captures the second face of R&D.

Cooperation

This explanatory variable comes from CIS. Firms are asked whether or not they cooperate with other firms with respect to their innovation activity. The variable we use is defined as

[122] Ideally, the global technological frontier is needed for our analysis to incorporate the idea of the distance to the frontier as potential determinant for higher productivity. The global frontier can be defined as the highest productivity level of an individual firm in the world. This definition is not feasible in practice, because we do not have worldwide micro data at our disposal.

the percentage of firms (as percentage of total number of firms) that reported cooperation.

Efficiency difference

As discussed in section 4.3.2, we want to test the importance of within industry variation. More precisely, we use the variance in average variable costs (variable costs over revenues) as indicator for differences in (cost) efficiency. Variable costs include wages and costs for intermediate inputs. If the variance is low, it points to small differences in performance across firms.

To some extent, this indicator is comparable to the variable that measures the distance to the frontier as they both measure differences within an industry. High values for both indicates large variation. But two distinctions are the following. First, this measure of efficiency difference uses the average variable costs, whereas the distance to the frontier is based on labor productivity. Second, if the distribution of the average variable costs is not normal (so not bell-shaped, with a peak at the mean) then these measures may provide different information. Relatively low variance can go together with a relatively large gap. Ignoring statistical outliers due to measurement issues, this means that most firms in this particular industry are relatively inefficient, while a limited number of firms are relatively efficient.

Funding

A government subsidy such as a R&D subsidy aims to stimulate innovation. Such subsidies reduce the innovation costs and help to internalize externalities. Our variable is based on the question in CIS whether or not the firm received a subsidy for its innovation activities. We use a ratio: the number of firms receiving a subsidy over the total number of firms (including non innovative firms) in the 3/4 digit SIC-code.[123]

Advertising costs

Advertisement expenses can form an entry barrier (see Sutton (1991)). For example, high advertisement expenditures may signal to potential entrants that they need a lot of advertisement to promote their products. However, high advertisement costs can also be a sign of intense competition in an industry (see *e.g.*, Creusen et al. (2006a)). Through advertising firms try

[123] We do not have (sufficient) data on the amount of innovation subsidies.

to make their products known to people, more transparent (*i.e.* promoting its features), so consumers will buy their product instead of products of their competitors. This indicator, expressed as ratio advertising costs over revenues, is derived from PS.

Cost disadvantage ratio

The cost disadvantage ratio is an indicator for entry barriers caused by economies of scale.[124] Economies of scale act as an entry barrier for new firms to enter the market if small firms have a cost disadvantage compared to big firms.

The cost disadvantage ratio in this chapter is defined as the ratio between the market shares of small and medium firms and the market shares of the large firms. More precisely, it is the ratio of the average labor productivity (defined as value added per worker) of the smallest firms responsible for 50 percent of the turnover in a market over the average labor productivity of the largest firms responsible for the remaining 50 percent of the turnover in a market. A low level of this ratio indicates economies of scale. This ratio comes from PS.

Turbulence

The turbulence indicator is defined as the number of firms that actually enter plus the number of firms that actually exit an industry related to the overall number of firms active in this industry. Although not necessarily directly related to competition, a high level of turbulence indicates that there are a lot of firms entering and/or leaving the market reflecting intense competition. This indicator is based on data from the General Business Register (ABR).

GDP

We use the change in real Gross Domestic Product (GDP) in the competition equation as a crude proxy for an increase in market demand.[125] The idea is that in a booming economy, demand (temporarily) exceeds supply. Then competing firms can set their prices above marginal cost and gain high profits without being impeded by competitors' price-cutting. Hence, excess demand is expected to weaken competition among firms. This GDP-measure (we use

[124] Nonetheless, prudence is called when using this indicator. Firms could also produce higher output in case of constant returns to scale because those firms are more efficient.

[125] We have no aggregate data for industry revenues per 3 digit industry. Moreover, such data probably enhances endogeneity issues more than using GDP.

the index) is based on data from the National Accounts of Statistics Netherlands.

Table 4.1 Descriptive statistics: Total economy

	Obs	Mean	Std. Dev.	Min	Max
Innovation intensity	1210	3.782	9.345	0.000	222.100
Competition (PE)	1179	5.341	3.781	– 1.140	38.823
Labor productivity	1179	84.300	108.436	10.860	1044.185
Efficiency difference	1179	0.018	0.012	0.001	0.085
Non-technological innovations	1210	0.410	0.202	0.000	1.000
Log capital intensity	1179	1.715	0.827	– 0.999	6.049
Number of employees	1179	17830	31217	45	249267
Turbulence	980	0.150	0.065	– 0.005	0.529
Advertising costs	1027	0.009	0.013	0.000	0.127
Disadvantage ratio	1179	0.788	0.919	– 14.640	18.879
GDP index	1207	131.553	7.286	115.103	142.276
Distance to frontier	1197	0.041	0.291	– 2.138	1.125
Cooperation	1210	0.163	0.144	0.000	0.811
Funding	1210	0.175	0.184	0.000	1.000

Note: Based on regression sample for equation (4.8).

Table 4.2 Descriptive statistics: Manufacturing

	Obs	Mean	Std. Dev.	Min	Max
Innovation intensity	745	5.430	11.520	0.009	222.100
Competition (PE)	738	6.632	3.934	– 0.043	38.823
Labor productivity	738	64.586	46.800	18.197	537.201
Efficiency difference	738	0.014	0.009	0.001	0.065
Non-technological innovations	745	0.475	0.195	0.000	1.000
Log capital intensity	738	1.896	0.681	– 0.435	4.650
Number of employees	738	6936	8237	105	55612
Turbulence	660	0.132	0.054	– 0.005	0.450
Advertising costs	666	0.009	0.012	0.000	0.125
Disadvantage ratio	738	0.856	0.933	– 6.293	18.879
Distance to frontier	743	0.014	0.300	– 2.138	1.125
Cooperation	745	0.216	0.150	0.000	0.811
Funding	745	0.259	0.186	0.000	1.000

Note: Based on regression sample for equation (4.8).

Table 4.3 Descriptive statistics: Services

	Obs	Mean	Std. Dev.	Min	Max
Innovation intensity	465	1.143	1.848	0.000	22.483
Competition (PE)	441	3.180	2.208	– 1.140	14.117
Labor productivity	441	117.290	161.459	10.860	1044.185
Efficiency difference	441	0.025	0.013	0.001	0.085
Non-technological innovations	465	0.306	0.168	0.000	0.842
Log capital intensity	441	1.414	0.952	– 0.999	6.049
Number of employees	441	36060	44307	45	249267
Turbulence	320	0.188	0.068	0.056	0.529
Advertising costs	361	0.008	0.014	0.000	0.127
Disadvantage ratio	441	0.675	0.885	– 14.640	5.716
Distance to frontier	454	0.085	0.272	– 0.819	0.903
Cooperation	465	0.076	0.079	0.000	0.528
Funding	465	0.041	0.057	0.000	0.421

Note: Based on regression sample for equation (4.8).

4.4.3 Descriptive statistics

Tables 4.1 to 4.3 show descriptive statistics of the key variables we want to use in section 4.5. We present figures for the total economy, but also for manufacturing and services. We distinguish between these two sectors as there might be differences between manufacturing and services. For instance, industries in manufacturing are often more exposed to foreign competition than industries in services (*e.g.* metal industry versus retail trade).[126] Indeed, Creusen et al. (2006b) find evidence for the Netherlands that competition is stronger in the manufacturing industry than in services. Additionally, the exact meaning of innovation activities is less clear in services than in manufacturing. For example, innovation in services tends to be organizational or client oriented rather than of a technological nature which is less difficult to define and measure.

Comparing the descriptive statistic of tables 4.1 – 4.3, there are substantial differences between both sectors. Despite the fact that the innovation intensity, level of competition, funding, cooperation and capital intensity are all higher in the manufacturing industry, labor productivity is higher in services. Differences in the remaining variables are less pronounced between the manufacturing industry and the services sector. With respect to labor productiv-

[126] Exceptions are transport and aviation.

ity and its key drivers, variances within the services sector are larger for labor productivity and capital intensity, but smaller for the innovation intensity and competition (according to the standard deviation).[127]

Finally, we correlate our indicator for efficiency differences (*i.e.* the variance of variable average costs) within an industry with the degree of competition to check whether the prediction of Aghion et al. (2005) is visible using simple correlations. They argue that the share of neck-and-neck industries (small variance of costs) will decline as competition increases. Translated to our situation, one would expect a positive correlation: more competition goes together with larger variances in variable costs suggesting less neck-and-neck industries. Figure 4.1 shows the result. We find a negative correlation, suggesting that the prediction of Aghion et al. (2005) is not right. A possible explanation for our finding is that as competition increases, inefficient firms leave the market and that reduces the variance. In the next section, we put the prediction of Aghion et al. (2005) to a further test taking account of other variables including industry and time fixed effects that might distort this correlation.

4.5 Empirical results

4.5.1 Explaining productivity

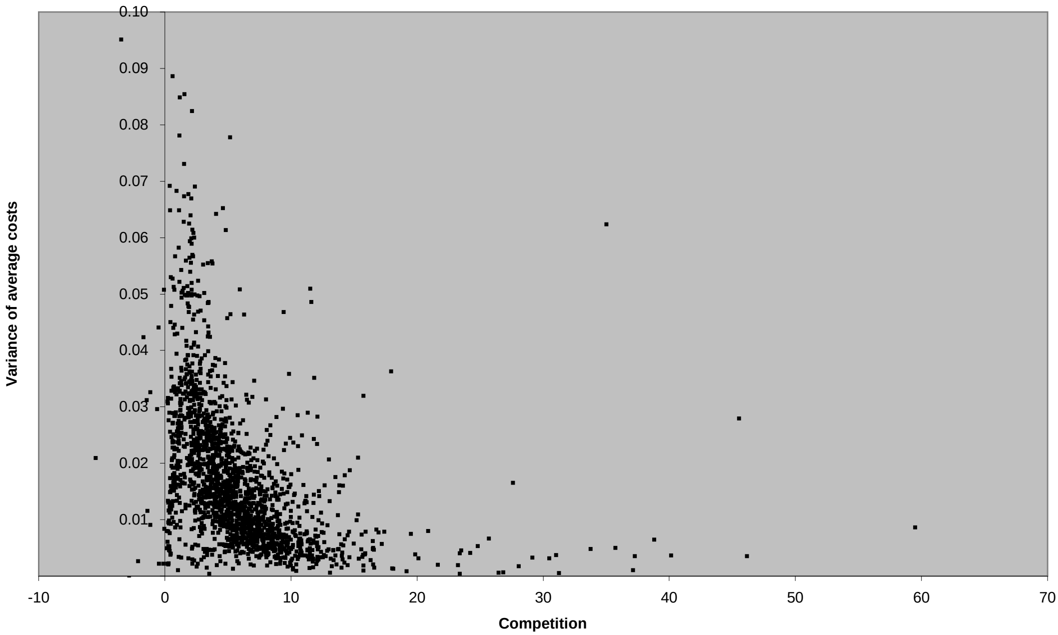
This section begins with addressing the research question to what extent competition and innovation enhance the productivity performance in the Netherlands.[128] Our starting point is equation (4.8) where we relate labor productivity to competition, innovation, distance to the frontier, capital intensity and economies of scale.[129] As discussed in section 4.3 other explanatory variables captured by the control vector X may also contribute to a better pro-

[127] We obtain negative results for turbulence, disadvantage ratio and distance to frontier. These observations do not affect the estimates because there are only a few of them. The explanation for these observations are as follows. The disadvantage ratio can be negative if the value added for an industry is negative. The distance to the frontier can become negative if the distribution of firms has a long tail where the average labor productivity is larger than that of the 75th percentile. The negative result for turbulence is due to data problems related to a lack of consistent time series.

[128] Results not reported in this section are available on request.

[129] As we estimate our equations in first differences unobserved industry heterogeneity is controlled for as long as this unobserved heterogeneity is constant over time.

Figure 4.1 Competition and efficiency differences



ductivity performance (*i.e.* higher TFP). Therefore, we tested other variables, but they were not significant. For instance, we added to our specification the following explanatory variables: non-technological innovations, the variance of efficiency, the interaction between non-technological innovation and innovation intensity, and the interaction between the variance in efficiency and innovation intensity.

Table 4.4 presents the econometric results with respect to the estimated labor productivity equation.[130] As discussed, we are particularly interested in the underlying sources of TFP. In general, these findings provide mixed evidence of explaining TFP-growth.

First, we discuss the impact of innovation and competition on productivity. Both explanatory variables seem to be drivers of productivity growth, at least for the total economy. Our empirical results confirm the assertion that competition may directly stimulate firms to at-

[130] The tables report two tests: the Hansen's J statistic and the GMM S statistics. The former tests the validity of the instruments used, and rejection implies that the instruments are not valid. We find p-values larger than 0.05 in all cases, so our instruments are both relevant and valid. The p-value of the GMM S statistics is in almost all cases larger than 0.05, implicating that we cannot reject the null hypothesis that our variables are exogenous. One exception is variant 1 in table 4.4. Finally, we use the robust standard errors to calculate the t-values in all tables.

Table 4.4 Labor productivity

Explanatory variables	(1) Total	(2) Total	(3) Total	(4) Manufact.	(5) Services
Competition (-1)	0.0129***	0.00979**	0.0106***	0.0141***	0.00942
	(3.14)	(2.23)	(2.85)	(4.04)	(0.94)
Innovation intensity (-1)	0.00358**	0.00229	0.00133	0.000909	0.00144
	(1.98)	(1.53)	(0.81)	(0.48)	(0.19)
Distance Frontier (-1)		– 0.0309	– 0.0242	– 0.0394	0.00325
		(-1.07)	(-0.57)	(-0.45)	(0.08)
Distance Frontier*comp (-1)			– 0.0266*	– 0.0390***	– 0.0469
			(-1.65)	(-2.72)	(-0.90)
Distance Frontier*innov (-1)			– 0.00341	– 0.00370	– 0.0129
			(-0.37)	(-0.28)	(-0.51)
Log capital intensity (-1)		0.291***	0.301***	0.274***	0.258***
		(9.47)	(9.72)	(8.46)	(5.25)
Economies of scale (-1)		– 0.0188	– 0.0124	0.0188	– 0.155***
		(-0.82)	(-0.54)	(0.70)	(-4.75)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No
Hansen's J	0.6969	0.5098	0.3198	0.3001	0.3107
GMM C statistic	0.0396	0.6585	0.7991	0.5345	0.7330
Observations	1005	759	759	498	261

Note: Robust z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The values reported for the Hansen's J test are the p-values for the null hypothesis of instrument validity. The values reported for GMM C (difference-in-Sargan test) statistics are the p-values for the null hypothesis whether variables are exogenous.

tain higher productivity levels by reducing X-inefficiencies and/or removing inefficient firms. This is the case in columns (1) to (4), but not for services. In addition, the general idea that innovation is an important driver behind productivity growth is supported as well (see column 1). However, this result is not statistically significant if we control for other regressors. This result is in line with Van der Wiel et al. (2008), who also did not find a significantly positive effect of R&D on the growth of TFP for the Netherlands. A reason for this finding could be that part of the (process) innovation is embodied in physical capital, already picked up by our capital intensity indicator.

Next, we do not find support for the view that the distance to the frontier itself acts as a driver of productivity at the industry level due to 'costless' technology transfers. Remarkably, the interaction of this variable with competition has a negative (but very limited) effect on productivity, suggesting the Schumpeterian effect dominates. More competition seems to induce firms to abstain from technology transfers from the global frontier. Apparently, the costs of imitating global technologies are too high and they cannot be recovered in times of fierce competition. Another explanation is that with intense competition there might be less spillovers available as firms (in other countries) are more inclined to keep their information secret.

As expected, capital intensity positively and significantly correlates with labor productivity in all cases. With regard to the existence of economies of scale, although not significant, the results are in line with what is found in the literature. There is one exception, the coefficient is negative and significantly different from zero for services, suggesting substantial decreasing economies of scale on average.[131]

To conclude, we find evidence for a positive effect of competition on productivity, whereas the positive coefficient for innovation is weakly significant. These are partial effects of competition and innovation. Which of these two sources eventually drives productivity is to be determined. We consider next the effect of competition on innovation. After that, we take into account that innovation can also influence competition.

4.5.2 Explaining innovation

We start with explaining innovation using equation (4.9). For the control variable W we include the following explanatory variables: distance to the frontier, cooperation, efficiency difference and funding. Table 4.5 shows the results for the total Dutch economy (see column 1-3), manufacturing (column 4) and services (column 5) respectively.

Starting with the results for the total Dutch economy, table 4.5 clearly illustrates that it matters whether one takes into account other explanatory variables for innovation. In terms of variables used, columns (1) and (2) are to a large extent directly comparable to the approach of Aghion et al. (2005), whereas column (3) shows the results of equation (4.9) when one

[131] This is in contrast with findings of Kox et al. (2010), but they limit their analysis to the (European) business services. Moreover, they differentiate within industries and find increasing returns to scale but mainly for small firms.

Table 4.5 Innovation

Explanatory variables	(1) Total	(2) Total	(3) Total	(4) Manufacturing	(5) Services
Competition (-1)	0.677*	– 0.113	0.198**	0.175*	– 0.120
	(1.88)	(-0.18)	(2.08)	(1.67)	(-0.64)
Competition (-1) squared		0.0271	– 0.00696**	– 0.00652*	0.0160
		(0.90)	(-2.31)	(-1.94)	(0.84)
Distance to frontier (-1)			0.180	0.291	– 0.314
			(0.59)	(0.74)	(-1.40)
Cooperation (-1)			0.978	0.565	0.449
			(0.57)	(0.30)	(0.27)
Efficiency difference (-1)			6.574	1.453	7.971*
			(0.73)	(0.12)	(1.62)
Funding (-1)			– 0.213	– 0.309	– 0.363
			(-0.19)	(-0.25)	(-0.14)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No
Hansen's J	0.9683	0.7138	0.6332	0.5850	0.9438
GMM C statistic	0.0911	0.5109	0.3808	0.5051	0.2750
Observations	1210	1210	822	558	264

Note: Robust z-statistics in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The values reported for the Hansen's J test are the p-values for the null hypothesis of instrument validity. The values reported for GMM C (difference-in-Sargan test) statistics are the p-values for the null hypothesis whether variables are exogenous.

also includes other explanatory variables that might affect innovation. Aghion et al. (2005) use only competition, competition squared and year effects in their regressions, but ignore the potential impact of other determinants.[132]

Ignoring non-linearity, we find a positive and significant impact of competition on innovation (see column (1)). Extending our analysis and taking account of non-linearity, we do not find significant evidence for an inverted U-curve relationship between competition and innovation for the Netherlands when neglecting other explanatory variables (see column (2)). However, conditional on those other variables, we come up with a statistically significant inverted U-curve (see column (3)). So, more competition initially leads to higher innova-

[132] They use other determinants but these determinants are only included as instruments for competition coping with the endogeneity problem (*i.e.* policy instruments (see section 4.3.3) and other instruments like import rate).

tion and then to lower innovation expenditures per employee.[133] Differentiating between manufacturing and services, it turns out that this macro finding rests upon the conditions in the manufacturing industry. There is no evidence for an inverted U-curve in Dutch services, whereas the outcome for the Dutch manufacturing is significant at the 10 percent significance level.[134] The latter corresponds with Aghion et al. (2005) as they only analyzed UK manufacturing.

The coefficients of the other explanatory variables are not significant, except for the indicator on differences in efficiency within an industry.[135] This indicator correlates positively and significantly with the innovation intensity for services at the 10 percent confidence level. So high variance in variable costs correlates with more innovation, while one would expect a negative correlation: firms innovate to escape their competitors that have more or less the same efficiency level. An explanation for this finding might be that lagging firms in services still have enough incentives to catch up due to low levels of competition in this sector of the economy. The descriptive statistics show that competition is less intense in services than in manufacturing. The textbox contains additional estimations where we consider a subsample of our data investigating the theoretical predication in Aghion et al. (2005) that the inverted U shape relationship between competition and innovation should be steeper in more neck-and-neck industries (*i.e.* industries with less variance in costs or industries closer to frontier). This prediction is, however, not consistent with our data.

The absence of significant determinants explaining innovation in services underlines the common view that innovation is hard to measure in services, even with the availability of CIS.

Interestingly from a policy perspective, we do not find evidence for a positive impact of an innovation subsidy to stimulate innovation. The coefficient of funding is not significant at any confidence interval. This finding suggests that if we distinguish between stimulating

[133] Notice that this result could be related to a different sample as the number of observations differs between column (2) and column (3) of table 4.5. We checked for this argument, but the inverted U-curve is still present if we run the variant of column (2) on the same sample, yet the shape is less steep.

[134] Our results for services are in line with Creusen et al. (2006c). They found no evidence for an inverted U in the Dutch retail trade.

[135] We tested also non technological innovations as additional control variable, but this variable was not significant either.

No steeper inverted U due to more neck-and-neckness

The table below shows the results for two variants based on a subsample of our data analyzing whether or not the inverted U between competition and innovation becomes steeper due to more neck-and-neckness as theoretically predicted by Aghion et al. (2005). This subsample includes industries with below median differences in average variable costs or below median distance to the frontier. It is assumed that those industries are more neck-and neck (or leveled) meaning less differences in applied technology exist between firms. Again, we find an inverted U for both variants, although not significant for the variant based on distance to the frontier (see column 3). Looking at the size of the coefficients of competition and competition squared, the theoretical prediction by Aghion et al. (2005) is, however, not supported because the inverted U relationship between competition and innovation is not steeper than that in our baseline results for the total economy (see column 1 versus column 2).

	(1)	(2)	(3)
Explanatory variables	Basic variant	Small variance	Close to frontier
Competition (-1)	0.198**	0.177*	0.224
	(2.08)	(1.70)	(1.26)
Competition (-1) squared	– 0.00696**	– 0.00750**	– 0.0117
	(-2.31)	(-1.99)	(-1.38)
Distance to frontier (-1)	0.180	1.014	1.236
	(0.59)	(1.43)	(1.28)
Cooperation (-1)	0.978	0.535	2.039
	(0.57)	(0.24)	(0.78)
Efficiency difference (-1)	6.574	15.08	5.490
	(0.73)	(0.96)	(0.45)
Funding (-1)	– 0.213	– 2.631**	– 1.714
	(-0.19)	(-2.56)	(-0.90)
Year dummies	Yes	Yes	Yes
Industry dummies	No	No	No
Observations	1090	399	450

innovation by either competition or by innovation subsidies then it makes more sense for policy makers to use the former as this policy option appears to be the most promising one.

Apparently, competition is the most important determinant of innovation and this determinant is not always conducive to the innovation expenditures. Taking the outcome for manufacturing, the competition intensity in terms of PE that maximizes the innovation level is approximately 13. This implicates that ten percent of the manufacturing industries operated at least one year beyond this maximal level in the period 1996-2006. When competition becomes too fierce it may therefore have a negative effect on productivity via lower innovation expenditures. However, combining the estimation results presented in table 4.4 and table 4.5, it turns out that this is at levels of competition that are far beyond levels observed in general.[136] Hence when it comes to productivity, more intense competition is always better. [137]

4.5.3 Explaining competition

After examining whether competition affects the size of innovation expenditures, this subsection investigates this causality the other way round by estimating equation (4.10).

The idea behind this alternative channel from innovation to competition is that (product) innovation leads to more product variety (or more product differentiation). This creates (new) niches in markets with lower intensity of competition as a consequence (see also Boone (2000b)). Or, high levels of innovation expenditures form an entry barrier reducing the degree of competitive pressure (see Sutton (1991)).

GDP, the disadvantage ratio, turbulence indicator (*i.e.* the ratio of death and birth of enterprizes over the number of active enterprizes in an industry) and advertising costs are included as control variables into equation (4.10).

Table 4.6 presents the results for five variants. For both the total economy (columns 1-3) and services (column 5), the empirical evidence for this feedback mechanism from innovation back to competition appears to be absent.[138] But in manufacturing, this mechanism is present and statistically significant at the 5 percent level. Hence, more innovation will initially lead to more intense competition, but beyond some level of innovation, more innovation

[136] For the total economy, this level of PE is almost 600, and for the manufacturing industry with 1200 even much higher.

[137] Estimation of the reduced form of the labor productivity equation supports these findings for a large range of PE values.

[138] It can be argued that longer time lags than one year are needed for this channel because innovation will not directly have implications for the intensity of competition. Due to the limited number of observations we cannot take more lags into consideration without losing significance.

Table 4.6 Competition

	(1)	(2)	(3)	(4)	(5)
Explanatory variables	Total	Total	Total	Manufacturing	Services
Innovation intensity (-1)	– 0.0671	– 0.0112	0.300	0.506**	– 0.358
	(-1.25)	(-0.12)	(1.62)	(2.50)	(-1.57)
Innovation intensity (-1) squared		– 0.000260	– 0.00209*	– 0.00341**	0.0126
		(-0.49)	(-1.72)	(-2.57)	(0.54)
GDP (-1)			– 0.00509	0.0216	– 0.0442
			(-0.07)	(0.22)	(-0.90)
Disadvantage ratio (-1)			– 0.505	– 0.618	0.174
			(-1.34)	(-1.59)	(1.37)
Turbulence (-1)			3.116	1.662	7.212**
			(0.79)	(0.36)	(2.44)
Advertising costs (-1)			– 32.52	– 29.48	– 30.52**
			(-1.47)	(-1.00)	(-2.18)
Year dummies	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No
Hansen's J	0.6013	0.6551	0.6942	0.6515	0.9336
GMM C statistic	0.3369	0.4171	0.5300	0.4348	0.0456
Observations	944	944	696	433	263

Note: Robust z-statistics in parentheses: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The values reported for the Hansen's J test are the p-values for the null hypothesis of instrument validity. The values reported for GMM C (difference-in-Sargan test) statistics are the p-values for the null hypothesis whether variables are exogenous.

expenditures will have a negative effect on measured competition in the Dutch manufacturing industry. Looking at our data set, this happens occasionally. It turns out that there were three manufacturing industries that were beyond this level in the period 1996-2006. These industries consist of a relatively small number of firms. So, these industries are concentrated and entry might be hindered due to high levels of innovation that act as entry barrier.

Looking at the other explanatory variables, the coefficient of advertising costs is negative and (weakly) significant for services.[139] Apparently, higher advertising costs reduce the

[139] We also tested an alternative indicator for measuring changes in conduct as explanatory variable for competition. Following Creusen et al. (2006a), we add a count dummy to the equation that is based on the knowledge that PE and PCM only differs in the interpretation of the change in competition in case of a change in conduct (see Boone et al. (2007a)). This additional dummy did not differ significantly from zero in the estimates.

intensity of competition as these costs are used as strategic weapon to lower product substitutability and to raise an entry barrier.[140] Finally, the coefficient of the turbulence indicator is positive and significant at the five percent level for services. Thus, higher turbulence correlates positively with competition. More entry or more firms leaving the market signals more intense competition.

To wrap up. We find (weak) evidence for the manufacturing industry that beyond some 'maximal' innovation level there may exist a negative feedback mechanism from innovation to the development of competition. However, these innovation levels are very high relative to mean values.

4.5.4 Robustness of results

Finally, this subsection focuses on the robustness of our findings using PCM as indicator for competition.[141] Our preferred measure for competition is PE. As discussed in Boone et al. (2007a) and chapter 2, this indicator is more robust with respect to the development of competition over time than other traditional indicators like PCM and concentration rates such as H.

At the industry level we find a negative correlation between PCM and PE as one would expect. Nevertheless, in approximately 40 percent of all observations, PCM suggests a rise (fall) in competition, whereas PE points to a fall (rise) in competition. Hence, these two competition measures differ and this may have consequences for our findings if we use PCM instead of PE. Besides measurement errors, this discrepancy between PCM and PE is due to the reallocation effect of market shares from inefficient to efficient firms (see Boone et al. (2007a)).

To check the robustness of our results for productivity and the inverted U curve, we use PCM as alternative indicator for competition instead of PE. Tables 4.7 and 4.8 report the results

[140] Remind that in theory, advertising costs may also have a positive impact on competition because it may increase market transparency.

[141] PCM is a better measure for competition than H as H will always be wrong in case (inefficient) firms are forced to leave the industry due to more aggressive behavior by firms (see Boone et al. (2007a)). This increase in H due to more intense competition goes against the traditional interpretation that a fall in H points to more intense competition.

Table 4.7 Labor productivity: Robustness check of PE as competition indicator

	(1)	(2)	(3)	(4)	(5)	(6)
Explanatory variable	Total	Manufact.	Services	Total	Manufact.	Services
Innovation intensity (-1)	0.00177	0.000907	0.00291	0.00125	0.000948	0.000602
	(1.07)	(0.49)	(0.41)	(0.77)	(0.51)	(0.08)
PE (-1)				0.00950**	0.0120***	0.00883
				(2.46)	(3.49)	(0.90)
PCM (-1)	– 0.389	– 0.752	– 0.240	– 0.0441	– 0.149	– 0.0267
	(-0.96)	(-1.39)	(-0.62)	(-0.14)	(-0.37)	(-0.08)
Distance Frontier	– 0.0116	0.0169	0.0196	– 0.0110	– 0.0185	– 0.000850
	(-0.26)	(0.18)	(0.49)	(-0.27)	(-0.21)	(-0.02)
Distance Frontier*comp	– 0.000863	– 0.00467	– 0.0355	– 0.0196	– 0.0242	– 0.0434
	(-0.04)	(-0.25)	(-0.70)	(-1.12)	(-1.42)	(-0.82)
Distance Frontier*innov	– 0.00548	– 0.0119	– 0.0265	– 0.00569	– 0.00551	– 0.0105
	(-0.55)	(-0.79)	(-1.13)	(-0.62)	(-0.40)	(-0.44)
Log capital intensity (-1)	0.289***	0.258***	0.233***	0.301***	0.268***	0.254***
	(8.39)	(6.98)	(5.15)	(9.17)	(7.89)	(5.19)
Economies of scale (-1)	– 0.0121	0.0217	-0.151***	– 0.0114	0.0215	– 0.155***
	(-0.53)	(0.84)	(-4.56)	(-0.50)	(0.83)	(-4.62)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Industry dummies	No	No	No	No	No	No
Hansen's J	0.6742	0.5807	0.5516	0.2760	0.1881	0.3450
GMM C statistic	0.9473	0.9881	0.2503	0.9708	0.8463	0.7771
Observations	759	498	261	759	498	261

Note: Robust z-statistics in parentheses: *** p$<$0.01, ** p$<$0.05, * p$<$0.1. The values reported for the Hansen's J test are the p-values for the null hypothesis of instrument validity. The values reported for GMM C (difference-in-Sargan test) statistics are the p-values for the null hypothesis whether variables are exogenous.

of this robustness check. Starting with the results for labor productivity in table 4.7, the coefficients for PCM are negative in almost all cases as one would expect (*i.e.* lower PCM signals fiercer competition and this has a positive effect on productivity). But none of these coefficients are significant at the ten percent significance level. Concerning innovation in table 4.8, the coefficients for PCM are not statistically significant either.

What do these findings mean? If only industry aggregate data are available, researchers can only use (industry) PCM as measure for competition since the estimation of PE needs

Table 4.8 Innovation: Robustness check of PE as competition indicator

	(1)	(2)	(3)	(4)	(5)
Explanatory variable	Total	Total	Total	Manufact	Services
PCM (-1)	– 1.676	– 6.132	– 21.47	– 22.91	– 7.697
	(-0.17)	(-0.19)	(-1.19)	(-0.86)	(-0.57)
PCM (-1) squared		12.40	31.63	65.15	7.362
		(0.22)	(1.22)	(1.04)	(0.47)
Distance to frontier (-1)			– 0.0329	0.112	– 0.177
			(-0.10)	(0.27)	(-0.60)
Cooperation (-1)			– 0.227	– 2.066	0.623
			(-0.13)	(-1.00)	(0.34)
Efficiency difference (-1)			– 1.361	– 40.22**	12.26*
			(-0.07)	(-2.22)	(1.72)
Funding (-1)			– 0.145	0.275	– 0.00618
			(-0.13)	(0.22)	(-0.00)
Hansen's J	0.7750	0.7823	0.6698	0.5877	0.9860
GMM C statistic	0.7219	0.7499	0.1307	0.1326	0.5343
Observations	1201	1201	820	558	262

Note: Robust z-statistics in parentheses: *** p<0.01, ** p<0.05, * p<0.1. The values reported for the Hansen's J test are the p-values for the null hypothesis of instrument validity. The values reported for GMM C (difference-in-Sargan test) statistics are the p-values for the null hypothesis whether variables are exogenous.

firm level data.[142] But, our results show that PCM is not a significant explanatory variable for either labor productivity or innovation intensity at the aggregate level, whereas PE is. This is even the case when we divide industries into high concentrated industries and low concentrated industries based on the median of the H. It can be argued that in relatively low concentrated industries PCM should perform better as proxy for competition than in high concentrated industries since the reallocation effects will be smaller, and hence the potential bias in PCM will be less (see Boone et al. (2007a) and chapter 2). The regression results (not reported) do not support this statement since PCM is in both cases not a significantly explanatory variable for innovation.

[142] The PCM measure can be derived from aggregate industry data on revenue and variable costs.

4.6 Concluding remarks

This chapter examines the relationship between competition, innovation and productivity for the Netherlands. In the theoretical and empirical literature there is no consensus on how competition affects innovation, and consequently productivity. Recent evidence suggests a non-linear relation between competition and innovation (see Aghion et al. (2005)) that might, therefore, have a negative impact on productivity when competition becomes too fierce. However, studies from Tingvall and Poldahl (2006), but also Aghion et al. (2005) itself, illustrate that the finding of an inverted-U shaped relation is sensitive to the choice of both the competition and innovation indicator.

We use industry level data for more than 150 3/4-digit SIC-industries based on aggregated Dutch firm level data covering almost the whole Dutch economy over the period 1996-2006. We employ the Profit Elasticity (PE) and innovation expenditures as indicators for competition and innovation respectively. The PE is a better measure than traditional indicators like concentration rates or price cost margins (PCM) for measuring competition (see *e.g.* Boone et al. (2007a)). Similarly, Brouwer (2007) claims that innovation expenditures are a better concept in this respect than for instance cited patents that were used in Aghion et al. (2005). Our model consists of three equations – labor productivity, innovation and competition – that are estimated using the Generalized Methods of Moments (GMM) and in that way coping with the endogeneity problem between competition and innovation.

The main findings of our analysis can be summarized as follows. First, we find strong evidence for a positive impact of competition on total factor productivity (TFP) at the industry level. Competition directly increases TFP by reducing X-inefficiencies and removing inefficient firms.

Second, this chapter finds evidence that there may exist an inverted U-curve between competition and innovation for the Netherlands, at least for manufacturing industries. This corresponds with findings of Aghion et al. (2005). Apparently, competition is the most important determinant of innovation but this determinant is not always conducive to innovation expenditures. When competition becomes too fierce it may therefore have a negative effect on productivity via lower innovation expenditures. However, combining all our estimation results, it turns out that this negative effect is at levels of competition that are far beyond levels observed in general.

Third, we find no evidence for a negative feedback mechanism from innovation back to competition for the aggregate economy. In the sense that high levels of innovation expenditures do not lead to lower competition intensity. For the manufacturing industry we do find indications for such a feedback, but this occurs at levels of innovation intensity that are hardly observed in our data set.

Lastly, as indicator for competition, we use the PE in this study. To test the robustness of this indicator, we also applied PCM as indicator. The latter turns out to be not significant in any equation concerning productivity or innovation, making the PE an interesting measure for future productivity research.

Our findings have implications for policy. Results reveal that the direct effect of more intense competition appears to increase productivity at the industry level in the Netherlands. But we also find that there may exist an inverted U-curve between competition and innovation. Consequently, there might be a trade-off between competition and innovation, and this has implications for policy makers. Yet, our combined results indicate that an indirect negative effect of competition on productivity through lower innovation expenditures arises only at very high levels of competition. Therefore, given current innovation policy intensifying competition is a promising option for policy makers to raise productivity: one of the main goals for policy in the Netherlands. Certainly, if we consider our findings for innovation policy. We do not find significant econometric evidence that our indicator for innovation subsidies positively affects innovation expenditures. The findings with respect to competition are in line with Kocsis et al. (2009) as they argue that an inverse U relationship between competition and innovation can go together with a positive effect of competition on productivity.

Summarizing the discussion, this chapter provides evidence for a new look at the inverted U-curve between competition and innovation as found by Aghion et al. (2005). We claim that when it comes to productivity, more intense competition is seemingly always better in the Dutch case. Chapter 5 further investigates the importance of more intense competition for enhancing innovation using firm level data.

5 Product innovation reduces competition intensity

5.1 Introduction

Firms innovate to raise their profits.[143] This can happen in a number of ways. The innovation can reduce the firm's production costs (this is usually called a process innovation) and hence may increase the firm's profits relative to its competitors that do not innovate. A firm's profit can also increase due to a product innovation (*i.e.* new good or services). In this case, the innovation differentiates a firm's product from its competitors' products. This differentiation can be horizontal (preferred by some consumers but not by all) or vertical (preferred by all consumers if sold at the same price).[144] With such product innovation, the firm is able to raise its product margin for two reasons. One is that consumers like the product better and hence are willing to pay more. The other reason is that by moving away from its competitors the firm gains market power (competitive pressure is reduced) which allows it to raise its price.

In this chapter we are interested in the effect of product differentiation related to making products less close substitutes, and hence less competitive. This effect is relevant for two reasons. The first is policy related. There is a debate whether there is a trade off between static and dynamic efficiency. The former relates to welfare with given technology (in the short term) and the latter relates to welfare due to innovations (in the long term). Traditionally, the answer depends on whether an increase in competition intensity leads to a fall or rise in innovation. Inspired by Schumpeters work (see *e.g.* Schumpeter (1934) and Schumpeter (1942)), positive as well as negative effects from competition on innovation can be found in theory and empirics, while recent literature points to an inverse U relation between competition and innovation: the positive effect dominates at low levels of competition and the negative effect at higher levels of competition (see Aghion et al. (2005)). The effect we are interested in is where firms innovate in order to reduce the intensity of competition. So again there is a trade off between competition and innovation but with the causality going from innovation to competition intensity (instead of the other way around). The second reason why we are interested

[143] This chapter is based on Boone et al. (2010b).

[144] To illustrate, everyone prefers a faster computer if sold at the same price (*i.e.* vertical product differentiation). However, some consumers prefer an Apple computer, others a Dell even when sold at exactly the same price (*i.e.* horizontal differentiation).

in this relation is research related. It follows from our analysis that competition becomes endogenous when firms innovate. Hence it needs to be instrumented when considering the relation between competition and innovation, especially in case of product innovation.

The way we identify the effect of product innovation reducing competition intensity is as follows. We use Dutch firm level data covering large parts of the Dutch economy over the period 1993-2006. Two data sources from Statistics Netherlands are merged containing figures to measure competition indicators and innovation indicators at the firm level. When we look at the association between competition and innovation, we find a positive correlation across all firms and industries. We find this positive correlation for two types of innovation indicators related to product innovation. The first type of innovation indicator captures whether a firm has applied for a patent. Since usually quite some time elapses between applying for a patent and introducing new products based on that patent, we conjecture that this variable is not affected by the endogeneity problem just described. That is, a firm that has applied for a patent is not (yet) able to use the patent to differentiate its products from its competitors, and hence affect the level of competition. The other type of indicator that we use is whether the firm has recently introduced new products in the market. If product differentiation plays a role, we expect to see an effect for innovation variables of this type on competition.

Indeed, once we look inside industries (by using industry or firm fixed effects), the correlation between competition and innovation remains positive for the variable based on patent applications but turns negative for variables capturing new products introduced in the market. That is, within a market (or industry) the firms that introduce new products are the ones that face relatively little competition. We interpret this as innovating firms differentiating themselves from competitors and in this way reducing the competitive pressure that they face in the market.

Summarizing we find that more intense competition stimulates an industry to innovate more, but within the industry the firms that have successfully introduced new products are the ones that face less intense competition.

This chapter is organized as follows. The next section discusses the (empirical) literature. Section 5.3 introduces a model that captures that more intense competition leads to more innovation in an industry. But within the industry, the firms that introduce new products are the ones that move away from competitors and in this way reduce the intensity of competition that they face. We also introduce appropriate measures for competition and innovation analyzing the connection between both measures. Data and variables for the empirical analysis are discussed in section 5.4. Section 5.5 explains our empirical strategy and shows some descriptive statistics. Section 5.6 presents our main results on the relationship between competition and product innovation and checks the robustness of these results. Section 5.7 summarizes and concludes.

5.2 Empirical literature

Firms are seeking profits (or rents) and innovation can give firms a (temporarily) monopoly position or a cost advantage over competitors. This allows them to have a higher mark up. A rise in competition may enhance a firm's incentives to improve efficiency or to innovate with the aim to protect or enlarge its market share. But, competition may also discourage innovation as the rewards to innovation are reduced (see *e.g.*, Griffith et al. (2006)). In fact, there are two general views on the relationship between competition and innovation.

First, the view of a positive effect of competition on innovation can be found in Schumpeter (1934) and Scherer (1980). The idea is that competition stimulates incumbents to innovate otherwise the firm is forced to leave the market. Aghion and Howitt (1999) formalize this mechanism in a theoretical model. More intense competition raises innovation activities by increasing the difference between post-innovation and pre-innovation rents.

Second, the view of a negative effect of competition on innovation originally stems from Schumpeter (1942). Fiercer competition generates less R&D, reducing the rate of innovation and hence economic growth. This is often called the 'Schumpeterian effect' of competition on innovation. The intuition is that because the expectation of high profits drives innovation, an increase in competition reduces innovation if it results in lower profits. Firms need (some) market power to limit access to their innovation and to provide the incentive to innovate. Using a Schumpeterian endogenous growth model, Aghion and Howitt (1992) show that an increase in product market competition has a negative effect on productivity growth by reducing the monopoly rents that reward innovation (see also Romer (1990) and Grossman and Helpman (1991)).

Recent work by Aghion et al. (2005) comes up with an inverted U relationship between competition and innovation. At a low level of competition, competition has a positive effect on innovation, whereas at a high level of competition, it reduces innovation.

Papers such as Geroski (1990); Nickell (1996), Blundell et al. (1995, 1999) and Carlin et al. (2004) that find a positive relation between competition and innovation, imply that there is no trade off between static and dynamic efficiency at all. But papers such as Aghion et al. (2005) that find a negative effect of competition on innovation imply an (ex ante) trade off. More competition leads to lower prices and hence enhances static efficiency. However, if more competition leads to less innovation, it reduces dynamic efficiency.

We document that there is a trade off between static and dynamic efficiency, but this one is ex post. Our results are consistent with the view that more competition makes industries more innovative. However, firms that have innovated manage (ex post) to reduce the competition intensity that they face (reverse causality). Thus we find ex post a trade off between dynamic and static efficiency. This finding has implications for policy makers as the message is different from an ex ante trade off: an increase in competition is always good for innovation.

We are not the first to point to the possibility of endogeneity of competition when firms innovate. Aghion et al. (2005) use policy instruments related to product market interventions that differ over industries and time to cope with this endogeneity problem.[145] In their paper, however, the estimated coefficients with and without instruments hardly differ. This can be seen as suggesting that endogeneity is not a problem. We show that when focusing on creating niches by product innovation, the endogeneity problem is present and actually quite severe.

5.3 Model and variables

5.3.1 Model

We use the following framework to analyze the relationship between competition intensity and innovation. Firms can produce either a standard version of the product or a differentiated version. We assume that standard products are closer substitutes than differentiated goods,

[145] Note that Aghion et al. (2005) discuss another endogeneity mechanism based on innovation as entry barrier. This comes from the work of Sutton (1991), which we discuss below.

but creating a differentiated version of a standard product requires an investment $\iota_i > 0$ from firm i.

Let n denote the total number of firms/products in the market with n_s the number of standard products and $n_d = n - n_s$ the number of differentiated goods. Firm i produces with efficiency level x_i. That is, the cost function to produce q units of output is given by $c(q,x_i) + f_i$ where f_i denotes the fixed cost of production and $c(0,x_i) = 0$. Further we assume $c_q > 0, c_{qq} \geq 0, c_x, c_{qx} < 0$

Given the distribution of standard and differentiated products, firm i with product choice $j \in \{s,d\}$ chooses output level q_i to maximize profits

$$\max_q \{p_i^j(q, q_{-i}^j, q_{-i}^{-j}, \theta)q - c(q,x_i)\} \tag{5.1}$$

We write the reduced form Nash equilibrium (variable) profits as

$$\begin{aligned}\pi_j(x_i,x_{-i},N_j,N_{-j},\theta) = {} & p_j(x_i,x_{-i},N_j,N_{-j},\theta)q_j(x_i,x_{-i},N_j,N_{-j},\theta) \\ & - c(q_j(x_i,x_{-i},N_j,N_{-j},\theta),x_i)\end{aligned} \tag{5.2}$$

where N_j (N_{-j}) denotes the set of firms with product choice j ($-j$), $p_j(.), q_j(.)$ denote the Nash equilibrium price and output function for a firm with product $j \in \{s,d\}$, and θ represents the intensity of competition.

We make the following assumptions on equilibrium price and output functions.

Assumption 1. *Equilibrium price and output satisfy*

$$\frac{d\left(\frac{p_j(x_i,x_{-i},N_j,N_{-j},\theta)q_j(x_i,x_{-i},N_j,N_{-j},\theta)}{c(q_j(x_i,x_{-i},N_j,N_{-j},\theta),x_i)}\right)}{dx_i} > 0$$

The expression $pq/c(q,x)$ captures the extent to which a firm with efficiency x can raise revenue above costs. If we have constant marginal costs $c(q,x) = c(x)q$, this expression is related to the price cost margin. With constant marginal costs, the assumption implies that more efficient firms have higher price cost margins. In standard IO models this is indeed the case. The assumption above generalizes this to other cost functions by assuming that more efficient firms have higher values for revenue over variable costs. Again in most IO models this is satisfied.

We make the following assumptions on reduced form profits.

Assumption 2. *Equilibrium profits satisfy*

$$\frac{d\pi_i}{dx_i} > 0$$

$$\frac{d^2 \ln \pi_i}{d\theta dx_i} > 0$$

$$\frac{d \ln \pi_s(x_i, x_{-i}, N_s, N_d, \theta)}{dx_i} > \frac{d \ln \pi_d(x_i, x_{-i}, N_s, N_d, \theta)}{dx_i}$$

$$\pi_d(x_i, x_{-i}, N_s \setminus \{i\}, N_d \cup \{i\}, \theta) - \pi_s(x_i, x_{-i}, N_s, N_d, \theta) > 0$$

$$\textit{for each } |N_s| > \bar{n}_s \textit{ for some } \bar{n}_s \geq 1$$

$$\frac{d\pi_d(x_i, x_{-i}, N_s \setminus \{i\}, N_d \cup \{i\}, \theta) - \pi_s(x_i, x_{-i}, N_s, N_d, \theta)}{d\theta} > 0$$

$$\frac{d\pi_d(x_i, x_{-i}, N_s \setminus \{i\}, N_d \cup \{i\}, \theta) - \pi_s(x_i, x_{-i}, N_s, N_d, \theta)}{d|N_d|} < 0$$

and for the least efficient firm l in each set N_s and N_d

$$\frac{d \ln \pi_j(x_l, x_{-l}, N_s, N_d, \theta)}{d\theta} < 0$$

where $l = \arg\min_{i \in N_j} x_i$ for $j = s, d$.

The first inequality implies that more efficient firms make higher profits. Second, the percentage increase in profits due to an increase in efficiency is higher in a more competitive market (higher θ). Put differently, in a more competitive market firms are punished more harshly for a fall in efficiency. Third, an increase in efficiency has a bigger impact on the profits of standard goods than on that of differentiated goods. That is, with a differentiated good a firm is "protected" from competitors and the same fall in efficiency has a smaller effect on profits compared to a firm producing the standard good. Fourth, if firm i moves from a standard to a differentiated product, it raises i's profits as long as there are at least $\bar{n}_s \geq 1$ firms left with a standard product. That is, if i is monopolist on the standard good market, it is not necessarily profitable to move to a differentiated product. Fifth, as the industry becomes more competitive, the incentive to differentiate increases. Sixth, the incentive to differentiate falls as (more) competitors differentiate. And, finally, as the industry becomes more competitive, the least efficient firm sees its profits decrease.

We use the following definition of an equilibrium.

Definition 1. *In equilibrium it is the case that sets N_s and N_d (with $|N_s| + |N_d| = n$ and $N_s \cap N_d = \emptyset$) satisfy the following conditions*

- *for each firm $i \in N_s$ it is the case that*

$$\pi_s(x_i, x_{-i}, N_s, N_d, \theta) \geq f_i$$
$$\pi_s(x_i, x_{-i}, N_s, N_d, \theta) > -\iota_i + \pi_d(x_i, x_{-i}, N_s \setminus \{i\}, N_d \cup \{i\}, \theta)$$

- *for each firm $i \in N_d$ it is the case that*

$$\pi_d(x_i, x_{-i}, N_s, N_d, \theta) \geq f_i + \iota_i$$
$$\pi_d(x_i, x_{-i}, N_s, N_d, \theta) > \iota_i + \pi_s(x_i, x_{-i}, N_s \cup \{i\}, N_d \setminus \{i\}, \theta)$$

Below we formalize the idea that more intense competition (higher θ) leads to more innovation. Let M denote the distribution function of ι_i. Then the probability that firm i innovates is given by

$$\Pr(i\ innovates) = M(\pi_d(x_i, x_{-i}, N_s, N_d, \theta) - \pi_s(x_i, x_{-i}, N_s \cup \{i\}, N_d \setminus \{i\}, \theta)) \tag{5.3}$$

Let G denote the distribution function of f_i. Then the probability that a firm producing a standard product is active in equilibrium is given by

$$\Pr(i\ enters) = G(\pi_s(x_i, x_{-i}, N_s, N_d, \theta)) \tag{5.4}$$

Let $G(f|\iota)$ denote the distribution function of f conditional on the innovation cost ι. Then the probability that a firm producing a differentiated product is active in equilibrium is given by

$$\Pr(i\ enters) = \int_0^{\pi_d(x_i, x_{-i}, N_s, N_d, \theta)} h(\iota) G(\pi_d(x_i, x_{-i}, N_s, N_d, \theta) - \iota | \iota) d\iota \tag{5.5}$$

Then we have the following result on the effect of competition on entry and innovation.

Proposition 1. *The probability that a firm i innovates increases with θ:*

$$\frac{d\Pr(i\ innovates)}{d\theta} > 0$$

The probability that the least efficient firm l enters decreases with θ:

$$\frac{d\Pr(l\ enters)}{d\theta} < 0$$

where $l = \arg\min_{i \in N_j} x_i$ for $j = s, d$.

The first result from proposition 1 follows from the assumption above that $\pi_d - \pi_s$ is increasing in θ. The second from the assumption that the least efficient firm's profits fall with θ. Hence as the competition intensity increases we expect fewer firms in the market and hence higher concentration.

The first result implies that as competition intensity increases more firms have an incentive to move away from their competitors by innovating and differentiating their product. Hence looking across industries, higher competition intensity correlates positively with innovation.

Together with the second result we see that more innovation (due to more intense competition) correlates with less entry into the industry. This is reminiscent of a result by Sutton (1991) where innovation expenditures create a barrier to entry and hence reduces the number of entrants. However, in Sutton's analysis firms cannot enter with a standard product (that does not require innovation expenditures). As new firms are forced to replicate the high innovation expenditure, entry becomes less profitable. In our analysis the higher competition intensity forces inefficient firms out of the market (or prevents them from entering). This is consistent with the finding in Boone et al. (2007a) that more intense competition is positively correlated with concentration (Sutton (1991)'s analysis implies a negative correlation: increasing innovation expenditures raises concentration and reduces competition intensity).

Looking within an industry, assumption 2 implies that the firms that have innovated and now sell new products, face less intense competition than the firms selling standard goods. Hence, controlling for industry or firm fixed effects, we expect that firms that sell new products face less intense competition. This is again different from Sutton (1991)'s result where all firms in an industry with high innovation expenditure face less intense competition (due to the entry barrier created by innovation expenditure).

5.3.2 Variables to implement the model

To test the above mentioned theory in the data, we need appropriate measures for competition and innovation.

For competition, we use the profit elasticity (PE), which was introduced by Boone (2008) and Boone et al. (2010a). As argued in these papers, the PE is a robust way to measure competition. Other well known competition measures like the Herfindahl (H) index and the price cost margin (PCM) are less suited for the current problem for the following reason.

First, as shown in proposition 1 an increase in competition intensity θ will reduce the number of firms in the market. This increase in H due to more intense competition goes against the "traditional" interpretation where a fall in H is seen as evidence of more intense competition.[146]

Second, in order to use firm fixed effects, we need a competition measure that varies between firms and correctly indicates competition. This is problematic with the PCM at the firm level. Firms that are less efficient will partly compensate their higher (marginal) costs by charging a higher price. However, it is unlikely that they can fully pass through their higher cost. Hence within an industry, firms with higher costs also have lower price cost margins. However, the standard interpretation is that lower PCM (here due to inefficiency) signals more intense competition. Industry average PCM is easier to interpret in terms of competition intensity,[147] but does not allow us to use industry fixed effects. Hence, we do not use PCM and H in this chapter.

The idea of the PE is to measure the slope of the relation between profits and efficiency. As stated in assumption 2, more intense competition makes the relation between profits and efficiency steeper. We use $PE_i = d\ln\pi_i/d\ln x_i > 0$ as a measure of competition. Section 5.5 explains how we exactly implement PE in the data.

Similarly, given our framework, we need two types of innovation measures in the context of the interaction with competition. One that indicates whether a firm has recently introduced new products in the market. If product differentiation plays a role, we expect to find an effect for innovation variables of this type on competition intensity. The second type of innovation indicator should also capture innovation but should not be affected by the endogeneity problem. So this indicator should not directly affect competition.

The Community Innovation Survey (CIS) can provide both type of indicators.[148] Here, we are mainly interested in the effect of competition on the output of the innovation process. As

[146] The reason for these seemingly contradictory effects of competition on H is the different way in which competition is intensified. The traditional interpretation looks at entry. More firms in the market lead to more competition and lower H. We consider a change in conduct, where more aggressive behavior by firms leads to the exit of less efficient firms and hence a rise in H.

[147] Although as shown by Boone et al. (2010a) this is not a perfect competition measure either, especially in concentrated industries. Indeed, chapter 4 shows that PCM is not a useful indicator for competition in explaining innovation.

[148] See section 5.4 for further discussion of this data source.

we want to disentangle whether innovation is used by firms to differentiate themselves from competitors and hence reduce competition. We, therefore, focus on indicators for product innovation as this type of innovation particularly enables firms to differentiate from their competitors by creating a niche. We use 'sales of products new to the firm', 'sales of products new to the market' and 'product innovation'. We also use 'applied for patents'. Since it takes time from applying for a patent to actually introducing a new product, it is unlikely that this indicator affects competition in this period.

We do not employ R&D expenditures (or its broader concept: innovation expenditure) because this indicator is not useful for the question we are interested in. The main problem with R&D expenditures is that the relation with output in terms of innovation is not clear. Moreover, it can be affected by competition in a way that is not relevant for us. When competition is low, there can be generous R&D budgets which are wasted on less relevant things. As competition intensity goes up, the budget may go down but R&D workers may work harder as they worry about their firm's survival. Hence as competition intensity goes up, R&D output increases (the effect we are interested in) while R&D inputs may fall. In contrast, our output indicators (except applied for patents) directly measure innovation in the form of market introduction of a new good or service.

Having the appropriate measures for competition and innovation we expect the following correlations in the data. As the intensity of competition θ increases in industry s the profit elasticity PE increases for firms in industry s. Moreover, in response to the increase in θ more firms innovate. Hence across industries we expect to find a positive correlation between innovation and PE. However, within an industry, innovation leads to a lower PE for innovating firms (see assumption 2) as these firms move away from their competitors. Most of our product innovation indicators consider implementation of innovations (like sales new to the firm or new to the industry). With indicators measuring the implementation of innovations, we expect a negative correlation between PE and product innovation once we control for industry or firm fixed effects. We also have a dummy measuring whether a firm applied for patents. This innovation measure does not imply that the innovation has already been implemented. Hence here we expect a positive correlation between PE and innovation even if we control for industry or firm fixed effects.

The following summarizes this discussion. Across industries an increase in PE is positively correlated with all our innovation measures. Once we control for industry or firm fixed effects, the correlation between PE and innovation is

- Positive for innovation indicators that do not imply the implementation of the product innovation and
- Negative for innovation indicators that measure the product innovations implemented by the firm.

5.4 Data

Here we outline the procedure followed to construct the panel data that we use. This panel data is obtained from matching two data sources from Statistics Netherlands – Production Surveys (PS) and CIS. Both sources are surveys from Statistics Netherlands and based on firm level data with the same unique identifier and a similar unit of observation (*i.e.* firm or enterprize). This unique identifier enables us to merge the two data sources at the firm level. The competition measure employed in this chapter is derived from PS, whereas the innovation measures stem from CIS.

5.4.1 Two sources: PS and CIS

The PS provide data on, amongst others, total sales, employment, value added and profitability on a yearly basis. Data from the PS is available for the years 1993 to 2006.[149] The PS is a sampled survey; only larger firms (*i.e.* more than 20 employees) are included in the sample each year. For smaller firms, sampling fractions decrease, and consequently most smaller firms will have gaps in the data for several years. Moreover, Statistics Netherlands apply a rotating sample method to reduce the administrative burden of (small) firms. This also reduces consecutive observations of firms.

In order to obtain reliable firm level data we performed several 'cleansing' activities. For instance, we removed: (i) observations of firms with no turnover and employment, (ii) the second observation of the same firm in one year, (iii) observation of year t+1 if a firm has identical output and employment data (or value added) in two consecutive years. In fact, we

[149] Except for transport and telecom, data for these industries covers the period 2000-2006.

use the same cleaned data set as applied in chapter 3.

Data on innovation activities for our study has been gathered from the Dutch section of the CIS. The CIS is a European harmonized questionnaire, held every two years, containing questions about innovative activities in enterprises.[150] This questionnaire follows the guidelines of the Oslo Manual for collecting innovation data (see OECD and Eurostat (1997) and OECD (2005)). CIS provides figures for the input, throughput and output of innovation activities. It focuses, amongst others, on innovation expenditures and (the effect) on process and product innovations.

Our innovation data covers the period 1996-2006. In fact, we use six consecutive waves of CIS for analyzing the dynamics of innovation in connection with competition: *i.e.* CIS2 for 1994-1996, CIS2,5 for 1996-1998, CIS3 for 1998-2000, CIS3,5 for 2000-2002, CIS4 for 2002-2004, and CIS4,5 for 2004-2006.[151] CIS only includes firms with at least ten or more employees, and samples firms with less than 50 employees.

The structure of all these CIS questionnaires is the same, but the (definition of the) variables included are not always identical. For example, both the definition of the indicator 'sales new to the market' and the indicator 'sales new to the firm' in 1996 are not completely the same as the definition of those indicators in 2006. In cooperation with Statistics Netherlands, we matched earlier variables as much as possible to the equivalent definition used in the last CIS.[152] Given our panel approach we only select variables from questions that are identical over time or are identical after some modification. Because of this we did not lose observations building our panel data set.

A main advantage of CIS is that after merging with PS one can directly relate firms' innovation activities to their performance in the product market.

[150] The CIS was launched by a number of countries as a result of the OECD initiative for setting up guidelines for innovation surveys. Those surveys emerged from a growing concern about the weaknesses of the traditional R&D surveys (see Van Leeuwen (2009)).

[151] We cannot use CIS1, covering the period 1992-1994, due to the use of different sampling frames. In contrast to other (Eurostat) countries, Statistics Netherlands has carried out two intervening surveys (*i.e.* CIS2,5 and CIS3,5).

[152] We are very grateful to Statistics Netherlands for their time consuming efforts to make the data of our CIS panel as much as possible comparable from 1996 to 2006. Given our knowledge of CIS, (one of the authors of this chapter was one of the founders of the CIS in the Netherlands) we were able to make this CIS panel richer by making more variables comparable over time.

CIS has several shortcomings that limit the options for research, particularly in terms of panel data (see also Van der Wiel (2001b)). We mention the most important ones.

First, the number of observations in CIS is low compared to that of PS due to a more limited sampling technique. This narrows the matching with the PS, considerably reducing the number of observations in our panel data. And small firms are therefore less represented in CIS than in PS, while those firms might be considered as important sources of innovativeness. Moreover, linking firm level data from CIS over time creates a loss of data as most firms are not surveyed in consecutive waves. Hence, we miss the complete innovation history for many firms complicating our analysis. Additionally, CIS contains firms belonging to industries that are not present in PS and vice versa. This reduces the number of industries that can be examined.

Second, CIS suffers from lower response rates than PS and the responses can be selective as it is likely that innovative firms are more inclined to respond than firms that do not innovate.

Third, the use of consecutive surveys of CIS may lead to double counting of some innovation activities, and consequently lead to overreporting innovation. Take for instance CIS3,5 and CIS4. The former covers the period 2000-2002, the latter covers the period 2002-2004. Hence, the year 2002 is twice present, that is, as end year of CIS3,5 and as starting year of CIS4. Consequently, firms that are innovative in 2002 may have reported that twice (*i.e.* both in CIS3,5 and CIS4). This can lead to double counting of innovation activities and hence measurement errors for variables that are based on this three year reference period. Since there is no further information available, we cannot say anything about the size of this measurement error.

Finally, CIS does not capture all relevant issues of innovation. For example, new firms entering the market are initially not included in the sample, while these firms may enter the market because they are innovative. The absence or low coverage of these starting firms may underestimate the importance of innovation. In our analysis we ignore these issues, but they should be kept in mind as caveats.

5.4.2 Variables

As discussed above, we are interested in whether or not firms differentiate their products in reaction to more intense competition. To identify this effect, we have chosen the following four innovation indicators from CIS that are closely related to product innovation.

Dummy applied for patent

This dummy indicates whether or not a firm applied for a patent during the relevant three year period. A patent in CIS is defined as a set of exclusive rights granted by a state (national government) to an inventor or their assignee for a limited period of time in exchange for a public disclosure of an invention.

This is the indicator that should not suffer (much) from endogeneity problems since there is quite some time between applying for a patent and implementing the patent in a successful product launch. This step between patent and product sold in the market usually takes on average 8 years, but with a lot of variation (see *e.g.* Ernst (2001)).[153] Since firms have only applied for patents, it is presumably too early for these innovations to affect revenues and hence PE as our indicator for competition.

With the following three innovation indicators we do expect an effect of product innovation on competition intensity.[154]

Dummy product innovation

A product innovation is defined as the market introduction of a new good or service or a significantly improved good or service with respect to its capabilities, components or subsystem during the relevant three years period. The innovation needs to be new to at least the firm, but it does not need to be new to its industry or market. In other words, it does not matter if the innovation was originally developed by the enterprize itself or by other enterprizes (*i.e.* imitation).

Dummy sales new to firm

This dummy variable takes value 1 for sales during the relevant three years period related to the introduction of a new or significantly improved product by a firm that was already supplied by some competitors on its market.

[153] Moreover, patent data are indicators of inventions, not necessarily leading to innovations.

[154] Note that the second and third indicator are a further distinction of the first one (dummy product innovation).

Dummy sales new to market

This dummy variable takes value 1 for sales during the relevant three years period related to the introduction of a new or significantly improved product by a firm that was new on its market.

We prefer to use dummy variables instead of, for example, the percentage of total turnover related to goods or services innovations new to the firm (or new to the market) for the following reasons. First, firms probably do not exactly know what the percentage is, but presumably they do know whether or not they have introduced new products. Second, the percentage sales of new products is itself affected by competition (if more competition leads to lower prices, it tends to raise sales), making the variable harder to interpret. Further, we prefer using the 'dummy product innovation' over the 'dummy process innovation' since our story is about product differentiation that may have an effect on the intensity of competition. Process innovation does not necessarily lead to reductions in competition.

Finally, we are aware that our indicators have weaknesses as well. With regard to patents, not every innovative firm applies for a patent due to, amongst others, high costs of application and a preference to keep the innovation secret. A disadvantage of this variable is also that firms have only applied for patents; they have not been granted (yet). Hence we do not know for sure whether the firms actually "invented" something new. Next, patents focus on new innovation activities ignoring the possibility of imitation that can also enhance firm's performance (see, inter alia, Griffith et al. (2004) for the so called second face of R&D expenditures).

5.5 Empirical strategy and descriptive statistics

5.5.1 Empirical strategy

The empirical analysis proceeds in two-steps. First, yearly PEs are estimated for each firm in the sample. Then the correlations between the estimated PEs and different innovation measures are estimated.

In order to estimate PE the first step specifies the following log linear functional form using firm fixed effect regression technique:

$$\ln \pi_{it} = \alpha_{st} + \beta_{st} \ln c_{it} + \gamma_{st} (\ln c_{it})^2 + \varepsilon_{it}$$

where $\pi_{it}$is the profit of firm i in year t, c_{it} captures its efficiency in year t and α_{st}, β_{st}, γ_{st} are industry-year specific parameters to be estimated.

In our application below we measure c_{it} as the firm's ratio between total variable costs and total revenues. By assumption 1 this variable is monotone in efficiency x_{it}. Given the functional form chosen above, the PE for firm i at time t is then calculated as:

$$PE_{it} = \beta_{st} + 2\gamma_{st} \ln c_{it}$$

so that across industries and years the variation in the estimated PE is due to both the estimates β_{st} and γ_{st} and to the variation in the marginal costs c_{it} while within industries and years variation in PE is given only by variation in the marginal cost c_{it}. We do not constrain estimation in such a way as to guarantee that the estimated PE_{it} are negative (meaning a higher costs implies lower profits, as theory predicts). Rather, as 16% of PE_{it} are estimated to be positive, we replace them with missing values before using them in the second step of the analysis.

In the second step the absolute value of the estimated PE is then used as explanatory variable in a linear regression where the dependent variable is a dummy variable measuring innovation, I_{it}. In fact the variable I_{it} takes the form of the four different dummy variables discussed in section 5.4.2, respectively measuring whether a firm "applied for a patent", "sold products new to the firm", "sold products new to the market" or "introduced a product innovation".

To estimate such correlations a simple OLS specification is first estimated

$$I_{it} = \alpha + \beta PE_{it} + \varepsilon_{it}$$

which is likely to suffer from endogeneity as on the one hand PE is a determinant of the decision of whether to innovate or not but on the other hand innovation itself might reduce competitive pressure on the innovating firms. Moreover, it is conceivable that industries differ in their level of innovation activity without a direct causal relationship with competition. This correlation then simply reflects other institutional features of the industry. One way to take up this problem is by using fixed effects that remove any spurious correlation or endogeneity. Therefore, we also estimate specifications with different fixed effects:

- year fixed effects: $I_{it} = \alpha_t + \beta PE_{it} + \varepsilon_{it}$
- industry and year fixed effects: $I_{it} = \alpha_s + \alpha_t + \beta PE_{it} + \varepsilon_{it}$
- firm fixed effects: $I_{it} = \alpha_i + \beta PE_{it} + \varepsilon_{it}$
- firm and year fixed effects: $I_{it} = \alpha_i + \alpha_t + \beta PE_{it} + \varepsilon_{it}$

The objective is to learn something about the structure of correlations in the data and the type of endogeneity affecting them. In all specifications, except those with firm fixed effects, the error term ε_{it} is clustered by firm in order to account for persistence in innovating behavior by firms.

To further investigate and check the robustness of the results we also use a logit specification.

We use this logit specification technique for the following reason. The outcome (response) variables of our product innovation measures are binary: 0 (no innovation) or 1 (innovation). As these indicators are binary the use of OLS-regression might be problematic because the assumptions of OLS are violated (*i.e.* OLS ignores the discreteness of the dependent variable and does not constrain predicted probabilities to be between zero and one). With OLS one assumes that innovation is linear in competition. A more appropriate model that handles binary variables better is the logit model with a cumulative logistic distribution function.[155] Using maximum likelihood estimation this model avoids events that occur with probability greater than one or less than zero. The probability of y=1 (or y=0) is P (or $1-P$) and will vary across individual firms as a function of the explanatory variables (X).

$$P_i = F(y_i) = \frac{e^{y_i}}{1+e^{y_i}}$$

with

$$y_i = log(\frac{P_i}{1-P_i}) = \beta_1 + \beta_2 X_{i1} + ...\beta_k X_{ik} + \mu_i$$

In terms of size, the coefficients of this estimation do not have a direct intuitive interpretation as, for instance, the marginal effect of a change in competition should be calculated. And this marginal effect is not constant because it depends on the value of X (in our case the intensity

[155] We also tried a probit specification. The choice between both models depends on the assumptions made about the error term. In case of the logit model, the cumulative distribution of μ_i is logistics, whereas μ_i is normal distributed in case of the probit model. Results on the sign and size of the coefficient are exactly these obtained with logit. We therefore limit our discussion here to the logit specification.

of competition). However, in this study, we are mainly interested in the sign and significance of the coefficients of the explanatory variable, and this sign is the same as the sign of the marginal effect in all cases.

5.5.2 Descriptive statistics

For estimating PE we have approximately 288 000 observations at our disposal for nearly 160 SIC 3-digit industries over the period 1993-2006. After matching the two data sources CIS and PS, we obtained around 18 000 to 26 000 observations (depending on the type of innovation indicator) for estimations of the relationship between innovation and competition.

Table 5.1 Summary statistics: step 1

Variable	Observations	Mean	Standard deviation
Profits	287971	1160.242	14888.74
Variable costs/revenues	287971	0.6538517	0.2968564
PE (absolute value)	238687	3.034485	4.337289

Table 5.1 reports summary statistics on profits (π_{it}) and our proxy for capturing efficiency (ratio variable costs/revenues (c_{it})) – the variables used in the estimation of the PE – and the result of the estimate of PE_{it} itself. There is wide variation in both profits and the level of competition as indicated by PE across industries. Table 5.2 reports summary statistics for the different innovation measures I_{it} and our estimates for PE_{it}. These variables are used in the second step where we estimate the relationship between innovation and competition. Note that the number of observations of PE_{it} is much smaller here than in table 5.1 as innovation measures are available only for even years and therefore almost half of the estimates PE_{it} are used in each regression.

Table 5.2 Summary statistics: step 2

Variable	Observations	Mean	Standard deviation
PE (absolute value)	119401	3.031459	4.295725
Applied for a patent	21184	0.0971488	0.2961673
Sold products new to the firm	26515	0.2360173	0.424641
Sold products new to the market	26515	0.1574203	0.3642035
Introduced a product innovation	18421	0.3486239	0.4765476

With regard to our product innovation indicators, the distribution of innovative activities are highly skewed, with the majority of firms reporting no activities in any year.[156] Looking more precisely, firms reported having introduced a product innovation the most, whereas the dummy applied for patent is reported the least on average across industries and time.[157] The latter is not that surprising as patents are not used by every firm or in every industry.

Table 5.3 Summary statistics: correlation matrix

	Newfrmd	Newmktd	Paap	Prodinn	PE
Sold products new to the firm	1.0000				
Sold products new to the market	0.4759	1.0000			
Applied for a patent	0.2822	0.3375	1.0000		
Introduced a product innovation	0.8893	0.6446	0.3279	1.0000	
PE	0.1361	0.1004	0.1081	0.1441	1.0000

Table 5.3 reports the correlation (coefficient) between our innovation indicators and competition. PE correlates positively with each measure of innovation, suggesting that more competition and more innovation goes together. The indicator 'applied for a patent' exhibits the least coherence with the other innovation indicators, suggesting that it picks up other innovative efforts as well.

5.6 Empirical results

5.6.1 Impact of competition on innovation

The OLS regressions reported in table 5.4 show a positive and significant correlation between PE and all the product innovation measures, meaning that higher competition is associated with higher innovative activities related to new or significantly improved products.

As explained above, the OLS estimates might be affected by endogeneity. If the endogeneity is due to the fact that innovation reduces competitive pressure on innovative firms, then the estimated coefficients in table 5.4 are upward biased estimates of the effect of com-

[156] Not reported, but there are also significant differences between industries.

[157] The sum of the means of products 'new to market' and 'new to the firm' is not exactly equal to the mean of 'introduced a product innovation', as the number of observations differ partly due to the absence of observations for product innovation in CIS4,5.

petitive pressure of innovation on productivity. The reason is as follows. For a given increase in innovation, the observed increase in competition is smaller than the underlying increase in competition (which is partly undone by firms differentiating themselves from competitors).

Table 5.4 OLS regressions: Innovation and competition

	(1)	(2)	(3)	(4)
	Applied patent	Product new to firm	Product new to market	Product innovation
PE	0.00636***	0.0162***	0.0117***	0.0160***
	(0.00073)	(0.0010)	(0.00095)	(0.0013)
Constant	0.0735***	0.178***	0.116***	0.284***
	(0.0033)	(0.0045)	(0.0040)	(0.0062)
Observations	18316	23191	23191	16425
R^2	0.01	0.03	0.02	0.03

Standard errors in parentheses are robust with respect to within firm correlation. *** p<0.01, ** p<0.05, * p<0.1

Surprisingly, at least at first sight, when we introduce industry or firm fixed effects, and irrespective of whether we also add year fixed effects or not, the findings are mostly reversed. Indeed, as reported in tables 5.5 and 5.6 (both without year fixed effects), a negative and significant correlation is found between PE and three out of four innovation measures: "sold a product new to the firm", "sold a product new to the market", "introduced a product innovation". Instead a positive correlation is still found between PE and the variable "applied for a patent".

Apparently only across industries higher competition seems to be associated with higher innovation, while within industries higher competition is associated with lower innovation. Yet the finding of a negative correlation between PE_{it} and I_{it} within industries can also be seen as evidence that in fact product innovation releases competitive pressure on the innovating firm.

Indeed negative correlations are found when the dependent variable is a variable measuring innovation introduced into the market, whereas a positive correlation is still found when the dependent variable is one that measures innovation not yet introduced into the market. One can argue that "applying for a patent" does not release competitive pressure on a firm as it has no effect on the output market yet, whereas “selling a product which is new to the firm" or "selling a product which is new to the market" or "introducing a product innovation" do

Table 5.5 Sector fixed effects regressions: Innovation and competition

	(1) Applied patent	(2) Product new to firm	(3) Product new to market	(4) Product innovation
PE	0.00272**	– 0.00490***	– 0.00282**	– 0.00455***
	(0.0011)	(0.0014)	(0.0011)	(0.0015)
Constant	0.0874***	0.255***	0.169***	0.362***
	(0.0044)	(0.0052)	(0.0040)	(0.0056)
Observations	18316	23191	23191	16425
Number of firms	12406	14753	14753	11441
R^2	0.090	0.201	0.154	0.202

Standard errors in parentheses are robust with respect to within firm correlation. *** p<0.01, ** p<0.05, * p<0.1

Table 5.6 Firm fixed effects regressions: Innovation and competition

	(1) Applied patent	(2) Product new to firm	(3) Product new to market	(4) Product innovation
PE	0.00188**	– 0.00550***	– 0.00250***	– 0.00413***
	(0.000742)	(0.000985)	(0.000817)	(0.00108)
Constant	– 0.0869**	0.0340***	0.0154***	0.000297***
	(0.0343)	(0.00609)	(0.00505)	(0.00007)
Observations	18316	23191	23191	16425
R^2	0.009	0.028	0.020	0.025

Standard errors in parentheses are robust with respect to within firm correlation. *** p<0.01, ** p<0.05, * p<0.1

directly affect competition in the market. The release in competitive pressure is therefore due to the innovating firm differentiating its product from those of its competitors in a market.

One could put forward an alternative story for innovating firms having lower PE levels due to better products. Assume that particularly efficient firms that innovate increase the quality of their products, and higher quality lead to higher marginal costs. Hence we get firms with high costs and relatively high profits (as they sell higher quality goods). This, however, does not lead to a flatter relation between costs and profits, but leads to a steeper slope and hence higher PE. In contrast, we find a negative effect of product innovation on PE.

5.6.2 Robustness

This section checks the robustness of the previous results to draw more reliable conclusions. We use the logit specification explained in section 5.5 to take into account that we use dummies that cannot be negative or larger than one.

Table 5.7 Logit regressions: Innovation and competition

	(1)	(2)	(3)	(4)
	Applied patent	Product new to firm	Product new to market	Product innovation
Without fixed effects				
PE	0.00455***	0.0158***	0.00970***	0.0184***
	(0.000535)	(0.000899)	(0.000749)	(0.00134)
Observations	18316	23191	23191	16425
Sector fixed effects				
PE	0.00123***	– 0.00432***	– 0.00147***	– 0.00448***
	(0.000460)	(0.000856)	(0.000501)	(0.00114)
Observations	17753	23106	23088	16354
Firm fixed effects				
PE	0.0319**	– 0.0283***	– 0.0199**	– 0.0304***
	(0.0139)	(0.00813)	(0.00866)	(0.0101)
Observations	1992	4738	3742	2815
Number of firms	715	1560	1207	1064

Standard errors in parentheses are robust with respect to within firm correlation. *** p<0.01, ** p<0.05, * p<0.1

Table 5.7 presents the results for using this logit specification estimation. The results are similar with respect to the signs to the previous ones and underlines our story that product innovation reduces the competitive pressure within industries. The coefficient for PE is positive for applied for patents whatever specific specification, whereas the sign of the coefficient of PE changes from positive to negative for the other innovation indicators when including industry or firm fixed effects in the logit estimation.[158]

[158] Again, not reported but findings are similar when time fixed effects are included.

5.7 Summary and conclusions

This chapter focuses on the effect of product differentiation related to making products less close substitutes, and hence less competitive. More intense competition leads industries to innovate more. However, firms innovate to reduce competition. Hence within an industry, successful innovators of new products are the ones that face less intense competition after the innovation.

We identify these effects of product innovation reducing competition intensity using Dutch firm level data covering large parts of the Dutch economy over the period 1993-2006. We exploit two types of innovation indicators to cope with reverse causality from innovation to the intensity of competition. One that measures whether or not an innovation is introduced into the market. The other type measures innovation not yet introduced into the market. We conjecture that this second type is not affected by the endogeneity problem. One can argue that the latter does not release competitive pressure on a firm as it has no effect on the output market yet, whereas the first type of innovation does directly affect competition in the market. The release in competitive pressure is therefore due to the innovating firm differentiating its product from those of its competitors in a market.

Our main findings are as follows. We come up with an alternative explanation for the negative correlation between competition and innovation, and hence for the trade off between static and dynamic efficiency. We claim, however, that the policy implication is the opposite: more competition is always better for (product) innovation in industries! However, firms that have innovated manage (ex post) to reduce the competition intensity that they face. Thus we find ex post a trade off between dynamic and static efficiency. Indeed, once we look inside industries (by using industry or firm fixed effects), the correlation between competition and innovation remains positive for the variable based on patent applications but turns negative for variables capturing new products introduced in the market. That is, within a market (or industry) the firms that introduce new products are the ones that face relatively little competition. We interpret this as innovating firms differentiating themselves from competitors and in this way reducing the competitive pressure that they face in the market.

6 Samenvatting (Summary in Dutch)

Waarom dit proefschrift?

Dit hoofdstuk geeft de Nederlandse samenvatting weer en beschrijft de redenen voor dit proefschrift. De rode draad in dit proefschrift is de relatie tussen concurrentie, innovatie en productiviteit. Het proefschrift bestaat uit vijf hoofdstukken waarin de resultaten staan beschreven van onderzoeken die gerelateerd zijn aan deze relatie.

Productiviteit is belangrijk voor een bedrijf en voor een land. Productiviteit geeft aan hoe efficiënt een bedrijf opereert. Productiever zijn vergeleken met je concurrent of vergeleken met vroeger betekent dat met dezelfde hoeveelheid inzet van mensen(arbeid) en machines (kapitaal) meer wordt geproduceerd, of anders gezegd, meer toegevoegde waarde wordt gecreëerd. Maar productiviteit is ook belangrijk voor een land. Het zegt iets over de levensstandaard, de welvaart, van mensen. Een hogere productiviteit wordt gezien als een belangrijke motor voor het realiseren van economische groei en daarmee van meer welvaart.

Op lange termijn wordt de economische groei bepaald door de toename van het arbeidsaanbod en de arbeidsproductiviteit. Gezien het teruglopende arbeidsaanbod in Nederland in de komende jaren wordt de toekomstige (duurzame) economische groei dus meer afhankelijk van de arbeidsproductiviteitstoename. Met het oog op de toekomst is verhoging van de arbeidsproductiviteit daardoor een belangrijke doelstelling van economisch beleid.

Het lijkt echter al tijden minder goed te gaan met de ontwikkeling van de arbeidsproductiviteit in Nederland. Nederland behoort sinds de jaren zeventig tot de landen met het hoogste arbeidsproductiviteitsniveau, maar met een relatief langzame groei sindsdien. Zowel in historisch als in internationaal perspectief blijft de groei achter.

Maar hoe krijg je een hogere productiviteit volgens de theorie? Grofweg zijn er twee theorieën over groei. Volgens de neo-klassieke groeitheorie komt de productiviteitsgroei door technologische vooruitgang die uit de lucht komt vallen als manna. Kortom, de groei komt vanzelf, bedrijven of beleidsmakers hoeven daar zelf niets voor te doen. De endogene groeitheorie zegt daarentegen dat de technologische vooruitgang wordt gedreven door concurrentie en innovatie.[159] In deze gedachte moet er wel wat door bedrijven of beleidsmakers worden gedaan. Door investeringen in onderzoek en ontwikkeling (R&D) worden nieuwe processen en producten ontwikkeld, innovatie dus, waarbij men voortborduurt op bestaande kennis. Innovatie levert een hogere productiviteit op. Procesinnovatie richt zich daarbij op nieuwe methoden en technieken om goedkoper te kunnen produceren en productiever te worden. Ook nieuwe, betere producten via productinnovatie leiden tot een hogere productiviteit, doordat ze meer waarde creëren per eenheid input. De mate van concurrentie speelt daarbij ook mee, zowel bij de beslissing om een bestaande technologie over te nemen als bij de ontwikkeling van nieuwe technologieën door R&D inspanningen. Bij beide type innovaties wordt de welvaart in de toekomst hoger (hogere dynamische efficiëntie).

Verschillen in productiviteitsniveaus tussen landen worden ook van belang gezien voor de groeimogelijkheden van de productiviteit. Als landen (of bedrijven) zich nog niet aan de grens van het technische kunnen bevinden, kunnen ze een snelle groei realiseren door bestaande, superieure technieken te imiteren van toplanden (of bedrijven). Hierbij is het uitvoeren van eigen R&D ook zinvol, want daardoor wordt het imiteren gemakkelijker omdat meer kennis beschikbaar is om kennis van anderen op te kunnen nemen. Daarentegen zijn de groeimogelijkheden beperkt voor landen die op de grens van het technische opereren, en is innovatie belangrijk. Nederland is zo'n land wat aan de technologische grens ligt, maar dat is gemiddeld genomen en geldt niet voor iedere bedrijfstak afzonderlijk.

Gegeven dat concurrentie en innovatie belangrijke determinanten zijn van productiviteit ligt het voor de hand dat beleidsmakers zich actief richten op beide om langs deze weg de productiviteit te bevorderen.

Maar wanneer en waarom zou de overheid zich met deze beide determinanten moeten bemoeien? Theoretisch gezien kan worden beargumenteerd dat marktwerking tot optimale welvaart leidt zolang er geen marktimperfecties zijn. Pas als marktimperfecties zich voordoen is er eventueel een actieve rol voor de overheid weggelegd. Een marktgebrek is bijvoorbeeld marktmacht. Door misbruik van marktmacht door bestaande bedrijven betalen consumenten te hoge prijzen, waardoor het gebruik van een bepaald product kleiner is dan wenselijk zou

[159] Dit proefschrift laat andere mogelijke determinanten van productiviteit buiten beschouwing. Zo kan menselijk kapitaal in de vorm van meer of beter onderwijs, of meer vaardigheden ook bijdragen aan een hogere productiviteit.

zijn. Bovendien kan marktmacht een reden zijn waarom bedrijven minder investeren en innoveren of dat er geen nieuwe concurrenten kunnen toetreden tot dezelfde markt.

Er zijn nog andere vormen van marktgebreken die redenen kunnen zijn voor overheidsbemoeienis. Als door positieve externe productie-effecten de (netto) maatschappelijke baten groter zijn dan de private baten, dan kan de hoeveelheid innovatieactiviteiten kleiner zijn dan maatschappelijk economisch wenselijk is. Enerzijds bestaan deze externe effecten uit positieve effecten van kennisspillovers (de voordelen van het voortbouwen op elkaar's schouders) en anderzijds uit negatieve externe effecten door duplicatie, afnemende meeropbrengsten (op elkaar's tenen staan) en de destructie van bestaande technologie als gevolg van de opkomst van nieuwe technologie (zogenaamde business stealing effect). Uit de empirische literatuur komt het beeld naar voren dat de maatschappelijke baten van innovatie veelal hoger zijn dan de private baten door externe effecten. Kortom, bij positieve externe effecten kan een bedrijf zich niet alle baten toe-eigenen van zijn inspanningen op gebied van product- en procesinnovaties, waardoor het minder prikkels heeft om aan innovatie te doen.

Beide hierboven beschreven marktimperfecties zijn dus mogelijke redenen voor de overheid om hierop beleid te voeren, waarbij de kosten van dat beleid (inclusief die voortkomen uit mogelijke overheidsgebreken) altijd kleiner moeten zijn dan de maatschappelijke baten. Ergo, het ingrijpen moet wel leiden tot meer welvaart. Hieronder volgt een kort overzicht van bestaand beleid rond concurrentie en innovatie.

Marktordening, mededinging en regulering staan hoog op de beleidsagenda. De afgelopen vijftien jaar zijn verschillende maatregelen genomen om voldoende marktwerking op productmarkten te bewerkstelligen. Voorbeelden op internationaal vlak zijn de éénwording van de interne markt van de Europese Unie in 1992, de Lissabon-agenda van de Europese Raad (2000) en verschillende overeenkomsten van de Wereld Handelsorganisatie. De Nederlandse overheid heeft met de nieuwe Mededingingwet van 1998 het mededingingsbeleid verder aangescherpt. Tevens werd de NMa opgericht om onder meer toezicht te houden op het handhaven van de concurrentieverhoudingen. Bovendien is de mededinging in specifieke markten aangepakt, hetzij door de hervorming en stroomlijning van regulering met de Marktwerking, Deregulering en Wetgevingskwaliteit (MDW-) operatie, hetzij door de privatisering van afgeschermde en gereguleerde markten zoals de telecommunicatiesector.

Het doel van innovatiebeleid is om R&D en daarmee innovaties in Nederland te bevorderen.

Een toename van R&D vergroot de welvaart, omdat R&D een hogere productiviteit geeft en positieve externe effecten met zich meebrengt. Programma's om private R&D activiteiten te stimuleren zijn o.a. WBSO (Wet Bevordering Stimulering Onderzoek). Daarnaast is er (Europees) beleid rond het vastleggen van intellectueel eigendomsrechten via patenten en octrooien. Zulke eigendomsrechten beschermen de innovator voor het weglekken van zijn/haar ideeën zonder een geldelijke vergoeding.

Wat heeft echter de theorie en empirie over concurrentie en innovatie te melden? Beide kunnen ieder op zich middelen zijn om de toekomstige productiviteit te verhogen, maar ze zijn ook onlosmakelijk met elkaar verbonden. Theoretisch gezien kan concurrentie langs twee kanalen de productiviteit bevorderen, waarbij in het tweede kanaal die verbondenheid tot uiting komt.

Ten eerste zal een sterkere concurrentie bedrijven prikkelen om efficiënter met bestaande middelen te produceren wat positief uitwerkt op productiviteit. Voldoende concurrentie en adequate wet- en regelgeving houden bedrijven scherp. Het dwingt ze om zo efficiënt mogelijk te werken. Merk op dat hier geen innovatie voor nodig is. Concurrentie zorgt er ook voor dat herstructureringsprocessen binnen een markt (of bedrijfstak) kunnen plaatsvinden, zodat efficiënte bedrijven marktaandeel winnen ten koste van inefficiënte bedrijven die zelfs van de markt kunnen worden gedrukt. Kortom, concurrentie zorgt via dit kanaal dat bestaande technieken efficiënter worden gebruikt. Dit leidt tot hogere productiviteit. In deze context wordt wel de term statische efficiëntie gebruikt. Een markt is statisch efficiënt wanneer een optimale combinatie van productiefactoren wordt gekozen binnen de restricties opgelegd door de bestaande technologie (productieve efficiëntie), en op basis daarvan een aansluiting van vraag en aanbod plaatsvindt die het totale surplus maximaliseert (allocatieve efficiëntie).

Een tweede, meer indirect, kanaal loopt via de invloed van concurrentie op innovatie. Concurrentie kan innovatie stimuleren, waarbij bedrijven door het op de markt zetten van nieuwe producten en/of betere productieprocessen hun concurrenten op achterstand trachten te zetten om zodoende meer winst te behalen (het zogenaamde 'escape competition'-effect). Als zo'n positief verband tussen concurrentie en innovatie bestaat, rijst de vraag welk beleid – concurrentie - of innovatiebeleid – effectiever en efficiënter is. Hiervoor is een kosten-batenanalyse nodig, waarbij de omvang van de externe effecten mede bepalend zal zijn.

Toch is de theorie en de empirie niet eenduidig over de relatie tussen concurrentie en innovatie. Ook een negatief verband is mogelijk en dat maakt de implicaties voor beleid gecompliceerder. Te veel concurrentie kan innovatie ook afschrikken, omdat het rendement op de innovatie naar verwachting te laag is door (te) sterke concurrentie (het zogenaamde Schumpeter-effect).

Recente theoretische inzichten ondersteunt met empirisch bewijs geeft aan dat de relatie tussen innovatie en concurrentie zelfs de vorm van een omgekeerde U kan aannemen (zie Aghion et al. (2005)). Deze vorm komt voort uit de aangehaalde tegengestelde effecten: escape effect versus Schumpeter effect. Deze recente literatuur gaat er vanuit dat als de concurrentie intensiever wordt, de innovatieactiviteit aanvankelijk eerst toeneemt. Als concurrentie heviger wordt, neemt de prikkel bij (efficiënte) bedrijven toe om zich te onderscheiden door een efficiëntievoordeel ten opzichte van concurrenten. Voor inefficiënte bedrijven nemen de prikkels om de achterstand in te halen juist af, omdat de kans om de achterstand nog in te lopen verdampt. Als de concurrentie op een bepaald moment te sterk wordt, daalt de innovatieactiviteit. Dus, bij verdere toename van de concurrentie gaat, vanaf een bepaald punt, het tweede effect domineren (namelijk Schumpeter effect), en leidt verdere verhoging van de concurrentie tot minder innovatie. Er is dus in deze gedachte sprake van een bepaald concurrentieniveau dat zorgt voor de hoogst mogelijk intensiteit van de innovatie. Voorbij dat punt is dus niet optimaal: het geeft minder innovatie!

Bij een negatief verband tussen concurrentie en innovatie ontstaat er mogelijk een afruil tussen statische en dynamische efficiency. Ook hier is een kosten-batenanalyse nodig voor het trekken van beleidsimplicaties. Belangrijke input (zoals uitgevoerd in dit proefschrift) hierbij is of zo'n negatief verband daadwerkelijk in de praktijk optreedt. Als dat het geval is, zullen beleidsmakers moeten kiezen tussen lage statische efficiency en hoge dynamische efficiency of omgekeerd, daarbij in acht nemend onzekerheden, risico's en voorkeuren van huidige en toekomstige generaties.

Bijdrage proefschrift

Dit proefschrift draagt bij aan de zoektocht naar de determinanten van productiviteit. Hierbij staat centraal hoe concurrentie en innovatie afzonderlijk, maar ook tezamen, de productiviteit kunnen bevorderen. Empirisch onderzoek naar de relatie tussen concurrentie en innovatie zoals Aghion et al. (2005) gedaan heeft voor het Verenigd Koninkrijk bestaat niet voor de Nederlandse economie. Dit boek gaat na of er een omgekeerde U bestaat voor Nederland en wanneer deze optreedt, zodat meer concurrentie ook kan leiden tot minder innovatie. Zolang hier geen beeld van is, kan het beleid niet zinvol van informatie worden voorzien welk beleid doelmatiger is: mededinging, innovatie of een combinatie. Dit is belangrijk want de Nederlandse prestaties op productiviteitsgebied in termen van groei zijn de laatste jaren mager in historische en internationaal perspectief. Mogelijk speelt hier een gebrek aan voldoende concurrentie.

De belangrijkste bijdragen van dit boek aan de bestaande literatuur naar de bronnen van productiviteit zijn de volgende vier punten.

Ten eerste zet het boek uiteen hoe concurrentie het best kan worden gemeten als men beschikt over bedrijfsgegevens. We gebruiken daarvoor de winstelasticiteit (PE) als nieuwe manier om concurrentie te meten. Deze indicator relateert de winsten van bedrijven aan hun efficiëntie. Hoe sterker de concurrentie hoe zwaarder inefficiënte bedrijven gestraft worden in termen van winst. We laten zien dat traditionele maatstaven als de Herfindahl index (H) en de prijskosten marge (PCM) een verandering in concurrentie niet altijd juist duiden (zie hoofdstukken 2 and 4). Deze verkeerde duiding treedt op als de intensiteit in concurrentie toeneemt doordat bedrijven onderling agressiever met elkaar zijn gaan concurreren. Dit gebeurt vooral bij bedrijfstakken/markten die geconcentreerd (c.q. weinig bedrijven) zijn en waarbij de verschuiving (lees reallocatie-effecten) van marktaandelen van inefficiënte bedrijven (met een lage PCM) naar efficiënte bedrijven (met een hoge PCM) groot is. Uit de analyse van de Nederlandse bedrijfsgegevens over de laatste vijftien jaar blijkt dat dergelijke verschillen in concurrentie duiding tussen PE en PCM regelmatig voorkomen. Voor toezichthouders op markwerking betekent dit dat ze voorzichtig moeten zijn met het gebruik van de PCM (en dat geldt zeker voor H) bij markten met een hoge concentratiegraad. En deze zijn nu net interessant vanuit mededingingsoogpunt. Hoofdstuk 3 laat zien hoe de PE in praktijk te meten is en dat deze indicator robuust is langs verschillende wegen.

Ten tweede onderzoekt het boek de relatie tussen concurrentie en innovatie (zie hoofdstukken 4 en 5). We doen soortgelijk onderzoek als Aghion et al. (2005) voor de Britse industrie heeft gedaan, maar nu voor de Nederlandse economie met bedrijfsgegevens, maar ook met bedrijfstakgegevens. Dit alleen al is informatief voor beleidsmakers, omdat zo'n analyse het inzicht in de relatie tussen concurrentie en innovatie vergroot. Immers, theorie en beschikbare empirie geven geen eenduidig antwoord hoe deze relatie eruit ziet. We nemen mee dat er een terugkoppelingsmechanisme kan zijn van innovatie naar concurrentie. Meer innovatie-uitgaven in de vorm van productdifferentiatie kan leiden tot niches met eventueel minder sterke concurrentie. We vinden aanwijzingen voor een omgekeerde U vorm tussen concurrentie en innovatie. Daarnaast zijn er aanwijzingen voor een terugkoppelingsmechanisme van innovatie naar concurrentie.

We gebruiken betere innovatie-indicatoren als in de studie van Aghion et al. (2005) om de relatie tussen concurrentie en innovatie te analyseren. In die studie wordt geciteerde patent data gebruikt als indicator voor innovatie, maar dit type data geeft onvoldoende alle mogelijke innovatieactiviteiten weer die kunnen leiden tot betere prestaties van een bedrijf. Veel innovaties worden niet gepatenteerd maar geheimgehouden. Wij nemen onder meer de innovatie-uitgaven van bedrijven als alternatieve maatstaf. Deze maatstaf omvat meer aspecten van innovatie dan patenten of R&D data.

Ten derde kijkt het boek ook naar de gevolgen van innovatie en concurrentie op productiviteit (zie hoofdstuk 4). Zowel innovatie als concurrentie moeten niet het doel van beleid zijn, omdat zij alleen maar middelen zijn die de productiviteit kunnen bevorderen. Hogere productiviteit (of meer welvaart) zou het beleidsdoel moeten zijn. Daarom is gekeken wat het effect is van innovatie en concurrentie op productiviteit. Het proefschrift bekijkt of meer concurrentie of stimulering van innovatie de productiviteitsprestaties in Nederland verbetert. Dit is interessant voor beleidsmakers gegeven de tegenvallende Nederlandse productiviteitsprestaties in internationaal en historisch perspectief. Onze bevindingen duiden er op dat meer marktwerking via een sterkere concurrentie een interessantere optie lijkt om de productiviteit te bevorderen dan stimulering van innovaties met subsidies. We laten echter ook zien dat er voor bepaalde bedrijfstakken wel degelijk een afruil kan bestaan tussen concurrentie en innovatie. Te sterke concurrentie is dan niet goed voor innovatie en daarmee voor de productiviteit. Deze bevindingen kunnen implicaties hebben voor de verdere invulling van het beleid rond marktwerking en innovatie. Echter, de empirie duidt er op dat zo'n afruil met negatieve gevolgen voor de productiviteit alleen optreedt bij uitzonderlijk hoge concurrentieniveaus. Kortom, in het algemeen is meer concurrentie goed voor productiviteit.

Tenslotte bevestigt dit boek het belang van bedrijfsgegevens voor onderzoek naar de relatie tussen concurrentie, innovatie en productiviteit, en het onderzoeken van bedrijfstakken buiten de industrie (zie hoofdstukken 2, 4 en 5). De beschikbaarheid van microdata geeft de mogelijkheid om rekening te houden met de heterogeniteit onder bedrijven. Verschillen in productiviteitsprestaties kunnen gerelateerd zijn aan de onderliggende determinanten van productiviteit. Bijvoorbeeld verschillen in toegepaste technieken, kwaliteit van het management, vaardigheden van het personeel en innovatie-inspanningen. Bovendien controleren we voor verschillen in instituties doordat we data hebben voor bijna de gehele Nederlandse economie. Verder relateren we bedrijfsgegevens met data op bedrijfstakniveau om rekening te houden met de variantie onder bedrijven en hun gemiddelde. Dit is een nieuwe manier om economisch gedrag te bestuderen met bedrijfstakgegevens. Tot slot, het onderzoek van Aghion et al. (2005) kijkt louter naar de Britse industrie. Dit proefschrift kijkt verder en neemt ook andere bedrijfstakken waaronder de dienstverlening onder de loep. We laten zien dat in Nederland de concurrentie in de industrie sterker is dan in de dienstensector. En dat we voor de industrie een omgekeerde U curve vinden tussen concurrentie en innovatie, maar niet voor de dienstensector.

De hoofdstukken kort belicht

De inhoud van de afzonderlijke hoofdstukken in dit proefschrift wordt hieronder kort belicht.

Hoofdstuk 1, *Introduction, framework and main results*, is een inleidend hoofdstuk. Veel wat daarin staat beschreven is hierboven al aan de orde geweest. Het hoofdstuk werkt de noties van de endogene groeitheorie verder uit naar een analytisch kader door verschillende theoretische en empirische studies tegen het licht te houden rond de onderlinge relaties tussen concurrentie, innovatie en productiviteit. Het beschrijft de achterliggende mechanismen en de link naar productiviteit. Bovendien geeft het mogelijke implicaties voor beleid weer. Ook gaat dit hoofdstuk als introductie verder in op wat concurrentie is en hoe het gemeten zou moeten worden. Het introduceert een nieuwe indicator voor concurrentie: de winstelasticiteit van een bedrijf (PE). Het centrale idee achter PE is dat in een competitieve markt inefficiënte bedrijven zwaarder worden gestraft in termen van winst. Een toename in (marginale) kosten van 1 procent leidt tot een sterkere winstdaling in een competitievere markt. Tenslotte bespreken we een aantal voor- en nadelen van diverse innovatie indicatoren.

Hoofdstuk 2, *Measuring Competition*, gaat dieper in op de nieuwe indicator voor concurrentie. De empirische IO literatuur gebruikt verschillende indicatoren voor meting van de concurrentie op een markt. H en PCM zijn het meest populair in deze literatuur. Echter, beide maatstaven hebben belangrijke bezwaren vanuit theoretisch perspectief die niet gelden voor de PE. Hoofdstuk 2 stelt de drie maatstaven op de proef langs twee wegen: (i) via simulaties en (ii) via data van bedrijven.

Daar er in de economische theorie geen duidelijke definitie bestaat van het concept "concurrentie", gaan we er van uit dat de concurrentie op een markt kan worden verhoogd op twee manieren. Ten eerste zal de concurrentie toenemen als er meer toetreders op de markt komen door minder toetredingsbelemmeringen. Ten tweede zal de concurrentie ook toenemen als bestaande bedrijven heftiger op elkaar gaan reageren door bijvoorbeeld het afschaffen van minimumprijzen of het doorbreken van een kartel. Beide manieren testen we uit via simulaties op de drie concurrentie indicatoren, met de vraag wat zeggen deze indicatoren over de ontwikkeling van de concurrentie.

Uit de simulaties blijkt dat ze alle drie bij minder toetredingsbelemmeringen duiden op een stijging van de concurrentie zoals de bedoeling is. Dit is niet het geval bij een verandering in het gedrag van bedrijven. In dit geval geeft H altijd een verkeerd signaal af over het verloop van concurrentie. Als bedrijven heftiger op elkaar reageren, neemt de concentratiegraad namelijk toe. Dit duidt op minder concurrentie, terwijl er juist sprake is van meer concurrentie. Dit verkeerde signaal wordt veroorzaakt doordat efficiënte bedrijven in deze situatie marktaandeel winnen (reallocatie effect) en/of heel zwak presterende bedrijven van de markt verdwijnen (selectie effect). Dus het aantal bedrijven in de markt neemt af en de scheefheid in de marktaandelen neemt toe, waardoor H stijgt.

Voor een deel van de simulaties geven PCM en PE ook hetzelfde signaal af over het verloop van de concurrentie als bedrijven heftiger op elkaar reageren. Dit geldt echter niet altijd. Ook de PCM kan afwijken van de PE door de gevolgen van de reallocatie en selectie effecten, terwijl de PE in grote mate consistent blijft in de duiding van concurrentie. Uit de simulaties blijkt dat hoe groter de omvang van het reallocatie effect hoe groter de kans is dat PCM omhoog gaat bij een toename van de concurrentie, en dus een verkeerd signaal afgeeft over de concurrentie. De vraag rijst hoe vaak dit in de praktijk voorkomt.

Gebruikmakend van CBS-bedrijfsgegevens uit de Produktie Statistieken (PS) over de periode 1993-2002, vergelijken we de PE met PCM voor 250 productmarkten in Nederland. We laten zien dat PE en PCM sterk zijn gecorreleerd, maar dat dat inderdaad niet altijd het geval is. In ongeveer één op de drie gevallen wijken ze van elkaar af. De verdere econometrische analyse bevestigt de simulatieresultaten dat de PCM een verkeerd signaal afgeeft over het verloop van concurrentie in markten met weinig bedrijven en hoge concentratiegraden. Merk op dat dit juist markten zijn die interessant zijn vanuit mededingingsoogpunt en regulering. Op grond van deze analyse concluderen we dat als men beschikt over bedrijfsgegevens de PE een meer betrouwbare meting geeft van concurrentie dan de PCM, laat staan H.

Hoofdstuk 3, *Robustness of Profit Elasticity*, gaat uitgebreid in op hoe de PE geschat kan worden in de praktijk als bedrijfsgegevens beschikbaar zijn. Ook wordt nagegaan hoe robuust de PE is op basis van Fixed Effect (FE) schatter vergeleken met alternatieve specificaties en schattingstechnieken. Speciale aandacht wordt daarbij geschonken aan hoe gevoelig de PE is voor de invloed van meetfouten en selectiviteitsproblemen. Om deze robuustheidchecks uit te voeren wordt een data set gebruikt met meer dan 320 000 observaties uit de PS over de periode 1993-2006. Deze data set bestaat uit circa 121 000 bedrijven uit 154 drie digit bedrijfstakken in Nederland. De belangrijkste resultaten van dit hoofdstuk zijn als volgt samen te vatten.

Ten eerste, onze voorkeur voor een loglineaire econometrische specificatie wordt ondersteund door testen. Een loglineaire vorm, waarbij zowel de winsten als de efficiency maatstaf in logs worden uitgedrukt, heeft als voordeel dat de geschatte coefficient voor PE direct interpreteerbaar is. Het is namelijk een elasticiteit: een hogere elasticiteit duidt op meer concurrentie. Het hoofdstuk laat via een aantal testen (o.a. Ramsey Reset test en de Box-Cox transformatie test) zien dat deze loglineaire specificatie niet wordt verworpen.

Ten tweede, het idee dat de relatie tussen winsten enerzijds en efficiëntie anderzijds iets zegt over concurrentie is robuust in de gebruikte schattingstechniek. Wij hebben de voorkeur om PE te schatten met FE, omdat deze schattingstechniek rekening houdt met zogenaamde bedrijfsspecifieke factoren. Bedrijven kunnen om allerlei, vaak niet observeerbare, redenen verschillen in hun prestaties. Door de FE-techniek schakelt men zoveel mogelijk die redenen uit door te veronderstellen dat ze constant zijn. Concreet vergelijken we de resultaten van een FE-model met die van OLS,'first differences', en Random Effect-model. Er blijkt een sterke significante correlatie tussen de uitkomsten van deze schattingstechnieken te zijn voor onze concurrentie indicator. Het bestaan van bedrijfsspecifieke factoren toetsen we verder door de karakteristieken van storingsterm van het model tegen het licht te houden. Is deze gecorreleerd in de tijd? Is deze dezelfde voor de bedrijven in het panel? Is deze gecorreleerd met een van de verklarende variabelen? Deze aanvullende econometrische toetsen (via o.a. F-testen en Hausman test (Hausman (1978))) ondersteunen onze voorkeur voor een FE model.

Ten derde, de resultaten voor PE blijken robuust te zijn als deze worden vergeleken met alternatieven die langs verschillende wegen rekening houden met mogelijke meetproblemen en selectiviteitsproblemen in de panel data set. We laten zien dat de PE sterk gecorreleerd is met deze alternatieven. Meetfouten in de verklarende variabele kunnen leiden tot een bias in de geschatte parameter voor de concurrentie-intensiteit. Alternatieve variabelen voor AVC, zoals de arbeidsproductiviteit, blijken sterk gecorreleerd te zijn met onze basisspecificatie, wat suggereert dat meetproblemen niet een doorslaggevende rol spelen in de bepaling van PE. Hetzelfde geldt voor mogelijke selectiviteitsproblemen. Omdat we gebruik maken van een steekproef, vallen door onder andere non-response en steekproefopzet bedrijven (tijdelijk) uit ons panel weg. Dit kan de schatting van PE vertekenen als dit wegvallen samenhangt met aspecten van concurrentie. Voor zover we kunnen nagaan, blijkt dit niet het geval te zijn als we rekening houden met de genoemde problemen.

Hoofdstuk 4, *Competition and innovation: Pushing productivity up or down?*, onderzoekt de relatie tussen concurrentie, innovatie en productiviteit. Zoals aan het begin van deze samenvatting is toegelicht zijn concurrentie en innovatie beide belangrijke determinanten voor productiviteit(sgroei). Productiviteitsgroei is op haar beurt een van de fundamentele drijvers voor een hogere welvaart door verbetering in de levensstandaard van mensen. In dat licht bezien is het niet vreemd dat het Nederlandse beleid zich richt op bevordering van de productiviteit via beleidsmaatregelen ter stimulering van marktwerking en innovaties.

In dit hoofdstuk houden we (nogmaals) de theorie rond concurrentie in relatie met innovatie en productiviteit tegen het licht. Het beeld is dan dat concurrentie normaal gesproken goed is voor productiviteit, als het gaat om het wegnemen van X-inefficiënties in het productieproces. Meer concurrentie leidt in de regel dan ook tot een hogere productiviteit en meer statische efficiëntie, omdat producenten gedwongen worden efficiënter te werken en inefficiënte bedrijven van de markt worden gedrukt. Theorie en empirie zijn minder eenduidig als de relatie tussen concurrentie en innovatie in dynamisch perspectief wordt bekeken. Zowel een positief als een negatief effect van concurrentie op innovatie zijn terug te vinden in de (empirische) literatuur uit de laatste decennia. Recente theoretische beschouwingen ondersteunt met empirische bevindingen van Aghion et al. (2005) voor de Britse industrie zetten de bestaande inzichten letterlijk op z'n kop met het vinden van een omgekeerde U vorm: meer concurrentie leidt eerst tot meer innovatie, maar voorbij een bepaald concurrentieniveau tot minder innovatie.

Om de relatie tussen concurrentie, innovatie en productiviteit voor Nederland te onderzoeken, wordt gebruik gemaakt van een gekoppelde dataset over de periode 1996-2006 uit de PS en de Community Innovation Survey (CIS). Van beide bronnen zijn de bedrijfsgegevens geaggregeerd tot veelal 3-digit bedrijfstakniveau en op dat niveau zijn vervolgens de innovatiegegevens gekoppeld met de productiviteitsprestaties. De dataset bestaat uit meer dan 150 bedrijfstakken die het overgrote deel van de Nederlandse economie omvatten.

Het vertrekpunt in de analyse is een productiefunctie waarin de gedachten van de endogene groei zijn opgenomen. Concurrentie en innovatie zijn hierin beide verklarende variabelen van de totale factor productiviteit (TFP). Bovendien voegen we uit de convergentieliteratuur het idee toe dat het uitmaakt voor de productiviteitstoename of bedrijfstakken wel of niet op de technologische grens zitten. Als niet dan zegt de afstand iets over de mogelijkheid van technologietransfers die extra productiviteitsgroei kunnen genereren. De afstand tot deze grens is bepaald met behulp van bedrijfsgegevens. We schatten uiteindelijk drie vergelijkingen – i.c. productiviteit, innovatie en concurrentie – waarbij we gebruikmaken van de GMM-schattingstechniek om rekening te houden met onderlinge endogeniteit en simultaniteit.

De belangrijkste conclusies uit dit hoofdstuk zijn de volgende. Alleen concurrentie blijkt een significant positief effect op de productiviteit te hebben als we controleren voor andere verklarende variabelen. Innovatie heeft wel een positief effect, maar is niet significant. We vinden echter ook aanwijzingen voor een omgekeerde U curve in Nederland. Net als Aghion et al. (2005) voor de Britse industrie vond, vinden wij dit voor de Nederlandse industrie maar niet voor de Nederlandse dienstensector. Deze omgekeerde U kan implicaties hebben voor het beleid, want te sterke concurrentie kan leiden tot minder innovatie-inspanningen en daarmee tot minder productiviteit. Echter, de regressieresultaten duiden er op dat deze negatieve gevolgen van concurrentie op de productiviteit door minder innovatie alleen optreden bij uitzonderlijk hoge concurrentieniveaus. Meer concurrentie is dus in de meeste gevallen goed voor de productiviteit. Meer concurrentie is dus een aantrekkelijk middel voor beleid om een hogere productiviteit te beogen zeker vergeleken met het stimuleren van innovatie met subsidies. Voor de laatste vinden we geen significant positief effect op innovatie. Net als in hoofdstuk 5 vinden we aanwijzingen dat de hoeveelheid innovatie-uitgaven ook een dempende werking kunnen hebben op de mate van concurrentie in de Nederlandse industrie: voorbij een bepaald niveau leidt meer innovatie tot minder concurrentie. Dit zogenaamde feedback mechanisme vinden we niet terug in andere Nederlandse bedrijfstakken.

Voor de mate van concurrentie gebruiken we de PE als indicator. Om de gevoeligheid van de uitkomsten te testen, kijken we ook naar de gevolgen van de inzet van PCM als alternatieve indicator voor concurrentie. De PCM blijkt echter geen significante bijdrage te leveren aan de verklaring van zowel de productiviteit als de innovatie. Ook pikt deze indicator geen informatie op als we hem naast de PE in de specificaties van productiviteit en innovatie opnemen.

Hoofdstuk 5, *Product innovation reduces competition intensity*, kijkt naar de relatie tussen concurrentie en (product)innovatie. Een manier voor bedrijven om via innovatie hun winstmarges te laten toenemen is door productdifferentiatie in de vorm van nieuwe goederen en diensten. De centrale hypothese in dit hoofdstuk is dan dat door zo'n productdifferentiatie de markt minder concurrerend kan worden. Enerzijds doordat bedrijven hogere winstmarges kunnen vragen omdat consumenten hun producten willen hebben en bereid zijn om er meer voor te willen betalen. Anderzijds vermindert de (gemeten) concurrentie doordat bedrijven op zoek gaan naar niches in de markt, wat hen marktmacht geeft met bijbehorende hogere marges. Dit mechanisme genereert ook een negatieve relatie tussen concurrentie en innovatie zoals Aghion et al. (2005) vond voor het Verenigd Koninkrijk, want concurrentie verhoogt in eerste instantie innovatie, maar door productinnovatie gaat de concurrentie weer omlaag. We tonen dit effect – waarbij concurrentie endogeen is geworden door de omgekeerde causaliteit met innovatie – aan en beweren dat de implicaties voor beleid anders zijn dan wat uit de afruil tussen concurrentie en innovatie volgt op basis van de bevindingen van Aghion et al. (2005). Dit doen we door gebruik te maken van gekoppelde bedrijfsgegevens uit de PS en de CIS over periode 1993-2006. Hiermee kunnen we zowel concurrentie als innovatie meten. Voor de meting van concurrentie gebruiken we de PE. Voor innovatie passen we twee type indicatoren toe. Het eerste type kijkt naar indicatoren waarvan we veronderstellen dat deze niet beïnvloed worden door het endogeniteitsprobleem tussen concurrentie en innovatie. We gebruiken hiervoor de indicator “een octrooi aangevraagd" door een bedrijf. In de regel zit er een behoorlijke tijd (oplopend tot gemiddeld wel 8 jaar) tussen deze aanvraag en het daadwerkelijk op de markt verschijnen van een nieuw product vallend onder dit octrooi. We veronderstellen dat dit type indicator (op korte termijn) niet de concurrentieverhoudingen verandert en positief correleert met PE. Het tweede type innovatie-indicatoren kijkt naar indicatoren die zich richten op recent gelanceerde productinnovaties. Hierbij is de gedachte dat als productdifferentiatie een rol speelt, we bij deze indicatoren een negatief effect verwachten op de mate van concurrentie. Dit hoofdstuk gebruikt hiervoor drie indicatoren rond productinnovatie uit CIS.

De belangrijkste bevindingen van dit hoofdstuk zijn als volgt samen te vatten. We komen met een alternatieve verklaring voor het bestaan van een negatieve correlatie tussen concurrentie en innovatie. De implicaties voor het beleid zijn echter anders. Aghion et al. (2005) laten zien dat bij te sterke concurrentie, een negatief effect optreedt op de hoeveelheid innovatie. Deze afruil tussen concurrentie en innovatie is hier ex ante. Onze bevindingen duiden echter op een afruil tussen beide die als ex post valt te interpreteren. Sterkere concurrentie leidt tot meer innovatie als over bedrijfstakken heen wordt gekeken, maar bedrijven binnen een bedrijfstak innoveren om de onderlinge concurrentie te verminderen. Binnen een bedrijfstak zijn het die bedrijven die succesvol nieuwe producten lanceren die minder concurrentie ondervinden. De resultaten zijn robuust voor toepassing van een andere schattingstechniek. De beleidsimplicatie die hieruit volgt is dat meer concurrentie altijd goed is voor productinnovatie, want we vinden geen negatief verband tussen concurrentie en innovatie.

Slotwoord

Hoewel hoofdstuk 5 en ook de eerdere hoofdstukken met een aantal bevindingen komen die een bijdrage leveren aan de verdere verklaring van productiviteit(sgroei), blijkt nog een groot aantal variabelen vanuit theoretisch oogpunt niet of nauwelijks econometrisch relevant te zijn in deze verklaring. Zo vinden we geen significante resultaten voor het belang van inhaalmogelijkheden naar de technologische grens. Enerzijds kan dit liggen aan dat de theorie over productiviteit nog onvolledig is. Anderzijds kan dit aan meetfouten liggen en dat benadrukt het belang van een goede meting door statistische bureaus. Er bestaat nog steeds een belangrijke kloof tussen theorie en meting van benodigde variabelen in de praktijk. In dit kader is bijvoorbeeld het goed waarnemen van menselijk kapitaal relevant. Deze factor vormt in de endogene groei een belangrijk fundament voor de verklaring van de groei van productiviteit, maar de statistische waarneming is nog onvoldoende om deze factor afdoende mee te nemen in productiviteitsanalyses op bedrijfsniveau. Ook datamateriaal dat helpt om een beter zicht te krijgen op de effecten van marktwerkings- en innovatiebeleid is nodig, maar ontbreekt op dit moment. Te denken valt aan onder meer NMa-interventies en veranderingen in innovatiebeleid. Deze zaken kunnen dienst doen als instrumenten om de endogeniteitsproblemen rond concurrentie en innovatie uitgebreider aan te pakken.

Tot slot. Dit proefschrift levert een bijdrage aan de zoektocht naar de bronnen van productiviteit die al lang geleden begonnen is. Het proefschrift geeft nieuwe inzichten of bevestigt inzichten die voor andere landen al bekend waren, maar nog niet voor Nederland. Veel vragen blijven echter nog onbeantwoord en daarom zal het onderzoek naar productiviteit zeker verder gaan. Een deel van de lastige achtbaan is doorlopen, maar onze hersenen (lees kennis) hebben nog niet alles goed kunnen ordenen wat er allemaal precies gebeurt rond productiviteit.

Curriculum Vitae

Henry van der Wiel was born in Zwijndrecht on 28 October 1962. After finalizing his Atheneum at Walburg-college in Zwijndrecht in 1981, he went to Erasmus University Rotterdam to study macro-economics. In 1988, he received his master in economics with the thesis *Het incidentele loon: het onberekenbare berekent*. Meanwhile, he joined the army to fulfil his military service in the period 1987-1988.

In 1988 he started to work as researcher at CPB Netherlands Bureau for Economic Policy Analysis. After working in a number of positions, he has been project manager of many projects since 2002. From early 2003 to mid 2004, he was associated with OCFEB Erasmus University Rotterdam for one day a week. From mid 2004 to present, he works as (PhD) researcher at CentER at Tilburg University for one day a week.

Van der Wiel has extensively done research and has frequently published on the sources of productivity growth at the macro-, meso- and firm level. He has been involved in several researches on the effects of ICT and innovation on the Dutch economy. Recently, he also took part in a number of researches that focussed on whether the degree of competition influences the performance of firms in the Netherlands. And looking for policy implications, he reviewed recent theoretical and empirical literature on competition and innovation issues.

Finally, he is member of various national advisory committees in the area of productivity, innovation and national accounts.

References

Abramovitz, M., 1956, Resources and output trends in the United States since 1870, *American Economic Review*, vol. 46, no. 2, pp. 5–23.

Acemoglu, D., P. Aghion and F. Zilibotti, 2006, Distance to frontier, selection and economic growth, *Journal of the European Economic Association*, vol. 4, no. 1, pp. 37–74.

Aghion, P., N. Bloom, R. Blundell, R. Griffith and P. Howitt, 2005, Competition and innovation: an inverted U relationship, *Quarterly Journal of Economics*, vol. CXX, no. 2, pp. 701–728.

Aghion, P., R. Blundell, R. Griffith, P. Howitt and S. Prantl, 2006, The effects of entry on incumbent innovation and productivity, Working Paper 12027, National Bureau of Economic Research, Cambridge MA.

Aghion, P. and P. Howitt, 1992, A model of growth through creative destruction, *Econometrica*, vol. 60, no. 2, pp. 323–351.

Aghion, P. and P. Howitt, 1999, *Endogenous Growth Theory*, The MIT Press, Cambridge.

Aghion, P., M. Dewatripont and P. Rey, 1999, Agency costs, firm behaviour and the nature of competition, CEPR Discussion Papers 2130, CEPR.

Aghion, P. and P. Howitt, 2006, Joseph schumpeter lecture appropriate growth policy: A unifying framework, *Journal of the European Economic Association*, vol. 4(2-3), pp. 269–314.

Amemiya, T., 1984, Tobit models: A survey, *Journal of Econometrics*, vol. 24 (1-2), pp. 3–61.

Amir, R., 2002, Market structure, scale economies and industry performance, Tech. rep., CORE.

Amir, R. and V. Lambson, 2000, On the effects of entry in Cournot markets, *Review of Economic Studies*, vol. 67, no. 2, pp. 235–254.

Ark, B. van, R. Inklaar and R. McGuckin, 2003, ICT and productivity in Europe and the United States: Where do the differences come from?, *CESifo Economic Studies*, vol. 49 (3), pp. 295–318.

Arrow, K., 1962, Economic welfare and the allocation of resources for innovations, in R. Nelson, ed., *The rate and direction of Inventive Activity*, Princeton University Press, Princeton.

Baily, M., C. Hulten and D. Campbell, 1992, Productivity dynamics in manufacturing plants, *Brookings Papers on Economic Activity*, vol. Microeconomics, no. 2, pp. 187–249.

Baldwin, J. and W. Gu, 2006, Competition, firm turnover and productivity growth, Economic Analysis Research Paper 42, Statistics Canada.

Balk, B., 2001, The residual: On monitoring and benchmarking firms, industries, and economies with respect to productivity, Tech. rep., ERIM, Erasmus University Rotterdam.

Bartelsman, E., J. Haltiwanger and S. Scarpetta, 2004, Microeconomic evidence of creative destruction in industrial and developing countries, Policy Research Working Paper 3464, World Bank.

Bartelsman, E., S. Scarpetta and F. Schivardi, 2003, Comparative analysis of firm demographics and survival: Micro-level evidence for the OECD countries, OECD Economics working paper 348, OECD.

Bartelsman, E.J. and M. Doms, 2000, Understanding productivity: Lessons from longitudinal microdata, *Journal of Economic Literature*, vol. 38, no. 3, pp. 569–594.

Baum, C., M. Schaffer and S. Stillman, 2003, Instrumental variables and GMM: Estimation and testing, Working Paper 545, Boston College.

Baumol, W., 2003, Four sources of innovation and stimulation of growth in the Dutch economy, Research Series 7, Ministry of Economic Affairs.

Berry, S., J. Levinsohn and A. Pakes, 1995, Automobile prices in market equilibrium, *Econometrica*, vol. 63, pp. 841–890.

Bikker, J. and M. van Leuvesteijn, 2005, An exploration into competition and efficiency in the Dutch life insurance industry, DNB Working Paper 47, DNB.

Bishop, S. and M. Walker, 2002, *The economics of EC competition law: concepts, applications and measurement*, Sweet & Maxwell.

Blanchard, O. and F. Giavazzi, 2003, Macroeconomic effects of regulation and deregulation in goods and labor markets, *Quarterly Journal of Economics*, vol. CXVIII, pp. 879–907.

Blundell, R., R. Griffith and J. van Reenen, 1995, Dynamic count data models of technological innovation, *Economic Journal*, vol. 105(429), pp. 333–344.

Blundell, R., R. Griffith and J. van Reenen, 1999, Market share, market value and innovation in a panel of British manufacturing firms, *Review of Economic Studies*, vol. 66(3), pp. 529–554.

Boone, J., 2000a, Competition, CEPR Discussion Paper 2636, CEPR.

Boone, J., 2000b, Competitive pressure: The effects on investments in product and process innovation, *RAND Journal of Economics*, vol. 31, no. 3, pp. 549–569.

Boone, J., 2001, Intensity of competition and the incentive to innovate, *International Journal of Industrial Organization*, vol. 19, pp. 705–726.

Boone, J., 2003, Optimal competition: a benchmark for competition policy, CentER DP 3, Tilburg University.

Boone, J., 2008, A new way to measure competition, *Economic Journal*, vol. 118, pp. 1245–1261.

Boone, J., R. Griffith and R. Harrison, 2005, Measuring competition, AIM Working Paper 22, IFS.

Boone, J., J. van Ours and H. van der Wiel, 2007a, How (not) to measure competition, TILEC Discussion Paper No. 2007-014, TILEC.

Boone, J., J. van Ours and H. van der Wiel, 2007b, How (not) to measure competition, CPB Discussion Paper 91, CPB.

Boone, J., J. van Ours and H. van der Wiel, 2010a, When is the price cost margin a safe way to measure changes in competition?, forthcoming.

Boone, J., E. Brouwer, L. Filistrucchi and H. van der Wiel, 2010b, Product innovation reduces competition intensity, forthcoming.

Boone, J. and E. van Damme, 2004, Markstructuur en innovatie, in B. Jacobs and J. Theeuwes, eds., *Innovatie in Nederland: De markt draalt en de overheid faalt*.

Box, G.E. and D. Cox, 1964, An analysis of transformations, *Journal of the Royal Statistical Society*, vol. 26, pp. 211–246.

Bresnahan, T. and P. Reiss, 1990, Entry in monopoly markets, *Review of Economic Studies*, vol. 57, pp. 531–553.

Bresnahan, T. and P. Reiss, 1991, Entry and competition in concentrated markets, *Journal of Political Economy*, vol. 99, no. 5, pp. 977–1009.

Brouwer, E., 1997, *Into innovation; Determinants and indicators*, Ph.D. thesis, University of Amsterdam.

Brouwer, E. and H. van der Wiel, 2010, Competition and innovation: Pushing productivity up or down?, forthcoming.

Brouwer, E., 2007, Innovatie en mededinging: Op zoek naar de bron van welvaart en vooruitgang, Discussion Paper 2007-021, TILEC.

Bulow, J. and P. Klemperer, 2002, Prices and the winner's curse, *RAND Journal of Economics*, vol. 33, no. 1, pp. 1–21.

Cabral, L., 2000, *Introduction to Industrial Organization*, MIT-Press, Cambridge Mass.

Cameron, A. and P.K. Trivedi, 2005, *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Cameron, G., 1998, *Innovation and Growth: a survey of the empirical evidence*, Ph.D. thesis, Nuffield College, Oxford.

Canton, E., 2002, Concurrentie en innovatie: Implicaties voor marktwerkingsbeleid: Literatuuroverzicht ten behoeve van de LT-studie van de afdeling kenniseconomie, CPB Memorandum 23, CPB.

Canton, E., B. Minne, A. Nieuwenhuis, B. Smid and M. van der Steeg, 2005, Human capital, R&D, and competition in macroeconomic analysis, CPB Document 91, CPB.

Carlin, W., M. Schaffer and P. Seabright, 2004, A minimum of rivalry: Evidence from transition economies on the importance of competition for innovation and growth, *Contributions to Economic Analysis and Policy*, vol. 3(1), pp. 1–43.

Cohen, W. and D. Levinthal, 1989, Innovation and learning: The two faces of R&D, *Economic Journal*, vol. 99, pp. 569–596.

Conway, P., D.D. Rosa, G. Nicoletti and F. Steiner, 2006, Regulation, competition and productivity convergence, Working Paper 509, OECD Economics Department.

Corts, K., 1999, Conduct parameters and the measurement of market power, *Journal of Econometrics*, vol. 88, pp. 227–250.

CPB, 2002, De pijlers onder de kenniseconomie: Opties voor institutionele vernieuwing, Bijzondere Publicatie 35, CPB.

Creusen, H., B. Minne and H. van der Wiel, 2006a, Competition in the Netherlands; an analysis of the period 1993-2001, CPB Document 136, CPB.

Creusen, H., B. Minne and H. van der Wiel, 2006b, Measuring competition in the Netherlands; a comparison of indicators over the period 1993-2001, CPB Memorandum 163, CPB.

Creusen, H., A. Meijer, G. Zwart and H. van der Wiel, 2008, Static efficiency in dutch supermarket chain, CPB Document 163, CPB.

Creusen, H., B. Vroomen, H. van der Wiel and A. Kuypers, 2006c, Dutch retail trade on the rise? Relation between competition, innovation and productivity, CPB Document 137, CPB.

Deelen, A., E. Jongen and S. Visser, 2006, Employment protection legislation; lessons from theoretical and empirical studies for the Dutch case, CPB Document 135, CPB.

Disney, R., J. Haskel and Y. Heden, 2000, Restructuring and productivity growth in UK manufacturing, Discussion paper 2463, Centre for Economic Policy Research.

Dixit, A. and J. Stiglitz, 1977, Monopolistic competition and optimum product diversity, *American Economic Review*, vol. 67, pp. 297–308.

Domowitz, I., G. Hubbard and B. Petersen, 1986, Business cycles and the relationship between concentration and price-cost margins, *Rand Journal of Economics*, vol. 17, no. 1, pp. 1–17.

Ernst, H., 2001, Patent applications and subsequent changes of performance: evidence from time-series cross-section analyses on the firm level, *Research Policy*, vol. 30, no. 1, pp. 143–157.

Fisher, F., 1987, On the misuse of the profits-sales ratio to infer monopoly power, *RAND Journal of Economics*, vol. 18, pp. 384–396.

Fisher, F. and J. McGowan, 1983, On the misuse of accounting rates of return to infer monopoly profits, *American Economic Review*, vol. 73, no. 1, pp. 82–97.

Gelauff, G., L. Klomp, S. Raes and T. Roelandt, eds., 2004, *Fostering Productivity: Patterns, Determinants and Policy Implications*, Elsevier Science Publishers.

Geroski, P., 1990, Innovation, technological opportunity and market structure, *Oxford Economic Papers*, vol. 42, pp. 586–602.

Geroski, P., 1995, What do we know about entry?, *International Journal of Industrial Organization*, vol. 13, pp. 421–440.

Gordon, R., 2000, Does the new economy measure up to the great inventions of the past?, *Journal of Economic Perspectives*, vol. 14, no. 4, pp. 49–77.

Griffith, R., S. Redding and J. van Reenen, 2004, Mapping the two faces of R&D: Productivity growth in a panel of OECD industries, *Review of Economics and Statistics*, vol. 86, no. 4, pp. 883–895.

Griffith, R., R. Harrison and H. Simpson, 2006, Product market reform and innovation in the EU, Working Paper 06/17, IFS.

Griliches, Z., 1998, *R&D and productivity: The econometric evidence*, Chicago University Press.

Griliches, Z. and J. Hausman, 1986, Errors in variables in panel data, *Journal of Econometrics*, vol. 31, pp. 98–118.

Grossman, E. and G. Helpman, 1991, *Innovation and growth in the global economy*, MIT Press, Cambridge, MA.

Hall, R., 1988, The relation between price and marginal cost in U.S. industry, *Journal of Political Economy*, vol. 96, no. 5, pp. 921–947.

Haltiwanger, J., 1997, Measuring and analyzing aggregate fluctuations: The importance of building from microeconomic evidence, *Economic Review*, vol. 79, pp. 55–77.

Hamberg, D., 1964, Size of firm, oligopoly, and research: The evidence, *Canadian Journal of Economics and Political Science*, vol. 30, pp. 62–75.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica*, vol. 50, no. 4, pp. 1029–1054.

Hausman, J., G. Leonard and J. Zona, 1994, Competitive analysis with differentiated products, *Annales D' Economie et de Statistique*, vol. 34, pp. 159–180.

Hausman, J.A. and D. Wise, 1977, Social experimentation, truncated distributions, and efficient estimation, *Econometrica, Econometric Society*, vol. 45(4), pp. 919–938.

Hausman, J.A. and D. Wise, 1979, Attrition bias in experimental and panel data: The Gary income maintenance experiment, *Econometrica, Econometric Society*, vol. 47(2), pp. 455–73.

Hausman, J.A., 1978, Specification tests in econometrics, *Econometrica*, vol. 46, no. issue 6, pp. 1251–71.

Heckman, J., 1979, Sample selection bias as a specification error, *Econometrica*, vol. 47, pp. 153–61.

Heckman, J., 1997, Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations, *Journal of Human Resources*, vol. 32(3), pp. 441–462.

Heckman, J., 2008, Econometric causality, Working paper 13934, National Bureau of Economic Research.

Heckman, J.J., R. LaLonde and J. Smith, 1999, The economics and econometrics of active labor market programs, in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics*, Amsterdam: North-Holland.

Hempell, T., G. van Leeuwen and H. van der Wiel, 2004, ICT, innovation and business performance in services: Evidence for Germany and the Netherlands, ZEW Discussion Papers 04-06, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research.

Hirano, K., G.W. Imbens, G. Ridder and D.B. Rubin, 2001, Combining panel data sets with attrition and refreshment samples, *Econometrica*, vol. 69, no. 6, pp. 1645–1659.

Hsiao, C., 2003, *Analysis of Panel data*, Cambridge University Press, New York, 2nd ed.

Jorgenson, D. and Z. Griliches, 1967, The explanation of productivity change, *Review of Economic Studies*, vol. 34, no. 99, pp. 249–280.

Jorgenson, D. and K.J. Stiroh, 2000, Raising the speed limit: U.S. economic growth in the information age, *Brookings Papers on Economic Activity*, pp. 125–211.

Kegels, C., M. Peneder and H. van der Wiel, 2008, Productivity performance in three small European countries: Austria, Belgium and the Netherlands, EU KLEMS Working Paper 21.

Kendrick, J.W., 1961, *Productivity Trends in the United States*, Princeton: Princeton University Press.

Kilponen, J. and T. Santavirta, 2007, When do R&D subsidies boost innovation? Revisiting the inverted U-shape, Discussion Paper 10, Bank of Finland.

Kleinknecht, A., K. van Montfoort and E. Brouwer, 2002, The non-trivial choice between innovation indicators, *Economics of Innovation and new Technology*, vol. 11(2), pp. 109–121.

Klette, T., 1999, Market power, scale economies and productivity: estimates from a panel of establishment data, *Journal of Industrial Economics*, vol. XLVII, no. 4, pp. 451–476.

Kocsis, V., R. Lukach, B. Minne, V. Shestalova, N. Zubanov and H. van der Wiel, 2009, Relation entry, exit and productivity: An overview of recent theoretical and empirical literature, CPB Document 180, CPB.

Kox, H., G. van Leeuwen and H. van der Wiel, 2010, Competitive, but too small! Productivity, regulation and entry-exit determinants in European business services, CPB Discussion Paper forthcoming, CPB.

Kraft, K., 1989, Market structure, firm characteristics and innovative activity, *The Journal of Industrial Economics*, vol. 37(3), pp. 329–336.

Leeuwen, G. van, 2009, *Innovation and performance; a collection of microdata studies*, Phd dissertation, Delft University of Technology.

Leeuwen, G. van, M. Polder and E. Veldhuizen, 2008, Micro-macro productiviteitsontwikkelingen: een case-studie voor de chemische industrie, Tech. rep., Statistics Netherlands.

Lever, M. and H. Nieuwenhuijsen, 1999, The impact of competition on productivity in Dutch manufacturing, in D. Audretsch and A. Thurik, eds., *Innovation, industry evolution, and employment*, Cambridge University Press.

Maddala, G.S., 2001, *An Introduction to Econometrics*, John Wiley & Sons.

Madsen, J.B., 2008, Semi-endogenous versus schumpeterian growth models: testing the knowledge production function using international data, *Journal of Economic Growth*, vol. 13, pp. 1–26.

Maliranta, M., M. Pajarinen, P. Rouvinen and P. Ylä-Anttila, 2007, Competition in Finland: Trends across business sectors in 1994-2004, KTM Julkaisuja 13, ETLA.

Mankiw, G. and M. Whinston, 1986, Free entry and social inefficiency, *Rand Journal of Economics*, vol. 17, pp. 48–58.

Mansfield, E., 1964, *Industrial Research and Technological Innovation: An Econometric Analysis*, New York, Norton.

Martin, S., 1993, Endogenous firm efficiency in a cournot principal-agent model, *Journal of Economic Theory*, vol. 59, pp. 445–450.

Mundlak, Y., 1978, On the pooling of time series and cross section data, *Econometrica*, vol. 46, pp. 69–85.

Nevo, A., 2001, Measuring market power in the ready-to-eat cereal industry, *Econometrica*, vol. 69, pp. 307–342.

Nickell, S., 1996, Competition and corporate performance, *Journal of Political Economy*, vol. 104, pp. 724–746.

OECD, 2005, *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data*, OECD, 3rd ed.

OECD and Eurostat, 1997, Oslo manual - proposed guidelines for collecting and interpreting technological innovation data, Tech. rep., OECD/Eurostat.

Oulton, N., 2001, ICT and productivity growth in the United Kingdom, Working Paper 140, Bank of England.

Panzar, J. and J. Rosse, 1987, Testing for “monopoly” equilibrium, *Journal of Industrial Economics*, vol. XXXV, pp. 443–456.

Patterson, K., 2000, *An Introduction to Applied Econometrics*, Macmillan.

Pilat, D. and F. Lee, 2001, Productivity growth in ICT-producing an ICT-using industries: A source of growth differentials in the OECD?, STI Working Paper 2001/4, OECD.

Porter, M., 1990, *The competitive advantage of nations*, MACMILLAN.

Ramsey, J.B., 1969, Tests for specification errors in classical linear least-squares regression analysis, *Journal of the Royal Statistical Society*, vol. 31, pp. 350–371.

Ridder, G., 1990, Attrition in multi-wave panel data, in J. Hartog, G. Ridder and J. Theeuwes, eds., *Panel Data and Labor Market Studies*, North-Holland, Amsterdam.

Roeger, W., 1995, Can imperfect competition explain the difference between primal and dual productivity measures? Estimates for U.S. manufacturing, *Journal of Political Economy*, vol. 103, no. 2, pp. 316–330.

Romer, P., 1990, Endogenous technological change, *Journal of Political Economy*, vol. 98, no. 5, pp. 71–102.

Salop, S.C., 1979, Monopolistic competition with outside goods, *Bell Journal of Economics, The Rand Corporation*, vol. 10(1), pp. 141–156.

Santarelli, E. and M. Vivarelli, 2006, Entrepreneurship and the process of firms' entry, survival and growth, Discussion Paper 2475, IZA.

Schaefer, K.C. and M.L. Visser, 2003, Reverse regression and orthogonal regression in employment discrimination analysis, *Journal of Forensic Economics*, vol. 16(3), pp. 283–298.

Scharfstein, D., 1988, Product market competition and managerial slack, *Rand Journal of Economics*, vol. 19, no. 1, pp. 147–155.

Scherer, F., 1967, Market structure and the employment of scientists and engineers, *American Economic Review*, vol. 57, pp. 524–531.

Scherer, F., 1980, *Industrial Market Structure and Economic Performance*, Rand McNally, Chicago, 2nd ed.

Scherer, F. and D. Ross, 1990, *Industrial market structure and economic performance*, Houghton Mifflin Company, 3rd ed.

Schumpeter, J., 1934, *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business*, Harvard University Press, Cambridge.

Schumpeter, J., 1942, *Capitalism, Socialism and Democracy*, New York: Harper.

Scott, J., 1984, Firm versus industry variability in R&D intensity, in Z. Griliches, ed., *R&D, Patents, and productivity*, pp. 233–248, Chicago: University of Chicago Press.

Sianesi, B. and J. van Reenen, 2003, The returns to education: Macroeconomics, *Journal of Economic Surveys*, vol. 17, no. 2, pp. 157–200.

Siebert, S. and N. Zubanov, 2008, Management economics in a large retail organization, IZA Discussion Paper 3645, IZA.

Solow, R., 1956, A contribution to the theory of economic growth, *Quarterly Journal of Economics*, vol. 70, no. 1, pp. 65–94.

Stiglitz, J., 1987, Imperfect information in the product market, in R. Schmalensee and R. Willig, eds., *Handbook of industrial organization*, vol. I, Elsevier Science Publishers.

Sutton, J., 1991, *Sunk Costs and Market Structure*, MIT Press, Cambridge, MA.

Symeonidis, G., 2001, Price competition, innovation and profitability: theory and UK evidence, CEPR Discussion Paper 2816, CEPR.

Tingvall, P.G. and A. Poldahl, 2006, Is there really an inverted U-shaped relation between competition and R&D?, *Economic Innovation New Technology*, vol. 15 (2), pp. 101–118.

Tirole, J., 1988, *The theory of industrial organization*, MIT Press.

Verbeek, M., 2004, *A Guide to Modern Econometrics*, John Wiley & Sons, ltd, second edition ed.

Verbeek, M. and T. Nijman, 1992, Testing for selectivity bias in panel data models, *International Economic Review*, vol. 33, pp. 681–703.

Vives, X., 2008, Innovation and competitive pressure, *The journal of Industrial Economics*, vol. 56, no. 3, pp. 419–469.

Wiel, H. van der, 2001a, Does ICT boost Dutch productivity growth?, CPB Document 16, CPB.

Wiel, H. van der, H. Creusen, G. van Leeuwen and E. van der Pijll, 2008, Cross your border and look around, CPB Document 170, CPB&Statistics Netherlands.

Wiel, H. van der and G. van Leeuwen, 2003, Do ICT spillovers matter? Evidence from Dutch firm-level data, CPB Discussion Paper 26, CPB.

Wiel, H. van der, 2001b, Innovation and productivity in services, *CPB Report*, vol. 1, pp. 29–36.

Wooldridge, J., 2002, *Econometric Analysis of Cross Section and Panel Data*, MIT Press, London.

Zhou, X., 2001, Understanding the determinants of managerial ownership and the link between ownership and performance: comment, *Journal of Financial Economics*, vol. 62, pp. 559–571.